University of Pennsylvania

The Wharton School

MGMT 111

Professor Adrian E. Tschoegl **Spring 1999**

Department of Management

Table of Contents

1. **Kobrin:** The MAI and the Clash of Globalizations. *Foreign Policy* 1998.
2. **Huntington:** The Clash of Civilizations. *Foreign Affairs* Summer 1993.
3. **Rapoport:** The importance of Space in Violent Ethno-religious strife. *Nationalism and Ethnic Politics* 1996
4. **Schwartz:** Why America Thinks It Has To Run The World *Atlantic* June 1996.
5. *Economist:* Cultural explanations: The man in the Baghdad Cafe. Nov 9, 1996.
6. **Griffin & Calabrese**: Coping With Extraterritoriality Disputes. *World Competition* 1988.
7. **McVey**: The Practical Side of International Trade. *The International Executive* 1989.
8. **Kindred**: Modern Methods of Processing Overseas Trade. *Journal of World Trade* 1988.
9. **Economist:** Schools Brief: Delivering the goods. 15 Nov 1997.
10. **Crowson**: Geographical shifts in the competitive strength of mineral production since 1960, and their causes. *Resources Policy* 1992.
11. **Weder:** How Domestic Demand Shapes the Pattern of International Trade. *World Economy* 1996
12. **Balasubramanyam & Balasubramanyam**: International Trade in Services: The Case of India's Computer Software. *World Economy* 1997.
13. **Barff & Austen**: 'It's gotta be da shoes': domestic manufacturing, international subcontracting, and the production of athletic footwear. *Environment and Planning A* 1993.
14. **Munk:** Carbitrage. *Forbes* Aug 14, 1995.
15. **Vernon**: International investment and international trade in the product cycle. *Quarterly Journal of Economics* 1966.
16. **Hayashi**: Restructuring of items within a product line: a case study of Japanese multinational enterprises in the electric machinery industry. *Environment and Planning A* 1994.
17. **Ho**: Industrial restructuring, the Singapore city-state, and the regional division of labor. *Environment and Planning A* 1994.
18. **Krugman:** Competitiveness: A Dangerous Obsession. *Foreign Policy* 1994
19. **Krugman:** Proving My Point. *Foreign Policy* 1994
20. **Lee & Swagel:** Trade Barriers and Trade Flows Across Countries and Industries. *Review of Economics and Statistics* 1997
21. **Rodrik:** Sense and Nonsense in the Globalization Debate. *Foreign Policy* 1997.
22. **Crook:** World trade: Jousting for advantage *The Economist* Sep 22, 1990.
23. **Lane:** World Trade: Time for another round. *The Economist* 3 Oct 1998.
24. **Godek**: Foreign Profits and the Political Economy of Quotas. *Cato Journal* 1991.
25. **Stiglitz:** Dumping on Free Trade: The U.S. Import Trade Laws. *Southern Economic Journal* 1997.
26. **Marsh:** Creating Barriers for Foreign Competitors: A Study of the Impact of Anti-Dumping Actions on the Performance of U.S. Firms. *Strategic Management Journal 1998.*
27. **Colgan:** Brave New World: International Regulation of Subsidies and the Future of State and Local Economic Development Programs *Economic Development Quarterly* May 1995.
28. **Breton & Salmon:** Are Discriminatory Procurement Policies Motivated by Protectionism? *Kyklos* 1996.

29. **Deardorf:** An Economist's Overview of the World Trade Organization. *Joint U.S.-Korea Academic Studies* 1997.
30. **Jackson:** The World Trade Organization: Watershed Innovation or Cautious Small Step Forward? *World Economy* 1995.
31. **Blackhurst:** The WTO and the Global Economy. *World Economy* 1997
32. **Arndt**: The Political Economy of Reciprocity. *Banca Nazionale del Lavoro Quarterly Review* Sep 1994.
33. **Nossal**: International sanctions as international punishment. *International Organization* 1989.
34. **Bayard & Elliot**: 'Aggressive Unilateralism' and Section 301: Market Opening or Market Closing. *World Economy* 1992.
35. **Bhagwati**: The Case for Free Trade. *Scientific American* 1993
36. **Daly**: The Perils of Free Trade. *Scientific American* 1993
37. **Barbier**: Managing Trade and Environment: The Demand for Raw Ivory in Japan and Hong Kong. *World Economy* 1991
38. **Khanna & Harford:** The ivory trade ban: Is it effective? *Ecological Economics* 1996
39. **Conybeare & Zinkula:** Who Voted Against the NAFTA? Trade Unions Versus Free Trade. *World Economy 1996.*
40. **Silvers:** Child Labor in Pakistan. *The Atlantic Monthly* Feb 1996.
41. **Hefeker:** Regional free trade and social integration *Public Choice* 1996.
42. **Baldwin:** The Causes of Regionalism. *World Economy* 1997
43. **Kobrin:** Regional integration in a globally networked economy. *Transnational Corporations* 1995.
44. ***Economist:*** Regionalism and trade: The right direction? Sep 16, 1995.
45. **Panagariya:** The Free Trade Area of the Americas: Good for Latin America? *World Economy* 1996
46. **Panagariya**: East Asia and the New Regionalism in World Trade. *World Economy* 1994
47. **Leipziger, Frischtak, Kharas and Normand:** MERCOSUR: Integration and Industrial Policy *World Economy* 1997
48. **Anderson & Smith:** NAFTA Expansion: US Imports Upon Chilean, Andean Pact, and MERCOSUR Accession
49. **Lustig.** Nafta: Setting the Record Straight. *World Economy* 1997.
50. **Gould:** Has NAFTA Changed North American Trade? *FRB Dallas Economic Review* 1998.
51. **McCarthy, Puffer & Simmonds**: Riding the Russian Roller Coaster: U.S. Firms' Recent Experience and Future Plans in the Former USSR. *California Management Review* Fall 1993.
52. **McCarthy & Puffer:** "Diamonds and Rust" on Russia's Road to Privatization: The Profits and Pitfalls for Western Managers. *Columbia Journal of World Business* Fall 1995
53. **Kramer:** Down on the Soviet Farm. *Technology Review* April 1996.
54. **Melnick & Kester**: The Foreign Exchange Market: Background Note and Problem Set. *Harvard Business School case 287-033* 1987.
55. **Dunn:** Retail Foreign Exchange Trading in Prague: Are Tourists Rational? *Journal of Soci-Economics* 1997
56. **Root**: The Balance of International Payments.
57. **Hopper:** What Determines the Exchange Rate: Economic Factors of Market Sentiment? *Federal Reserve Bank of Philadelphia Business Review* Sep/Oct 1997.
58. **Humphrey**: Explaining Exchange Rate Behavior: An Augmented Version of the Monetary Approach. Federal Reserve Bank of Richmond *Economic Review* 1981.
59. **Tschoegl**: The Yen and Purchasing Power Parity. SBCI Securities 1988.
60. ***Economist:*** Big MacCurrencies. 11 Apr 1998.
61. **Schwartz:** U.S. Foreign Exchange Market Intervention Since 1962. *Scottish Journal of Political Economy* September 1996.
62. **Howell**: The Role of the Bank for International Settlements in Central Bank Cooperation. *Journal of European Economic History* 1993.
63. ***Economist:*** Can the G-7 ride again? June 22, 1996.

64. *Economist:* Exchange rates: The fixed and the flexible. March 16, 1996.
65. **Chandler**: The Evolution of Modern Global Competition.
66. **Woolridge**: Multinational. *The Economist* June 24th 1995
67. **Kogut**: Foreign Direct Investment as a Sequential Process. C.P. Kindleberger and D. Audretsch, eds., *Multinational Corporations in the 1980s* 1983.
68. **Hu:** Global or Stateless Corporations Are National Firms with International Operations. *California Management Review* 1992
69. **Rimmer**: The internationalization of engineering consultancies problems of breaking into the club. *Environment and Planning A* 1988.
70. **Spar:** Lawyers Abroad: The Internationalization of Legal Practice. *California Management Review* 1997.
71. **Stopford & Strange**: The New Diplomacy. *Rival States, Rival Firms* 1991.
72. **Porter**: The Competitive Advantage of Nations. *Harvard Business Review* 1990.
73. **Kogut**: Country Capabilities and the Permeability of Borders. *Strategic Management Journal* 1991.
74. **Dunning:** The geographical sources of the competitiveness of firms: some results of a new survey. *Transnational Corporations* 1996.

WHARTON REPROGRAPHICS

The MAI and the Clash of Globalizations

by Stephen J. Kobrin

The Preamble Collaborative is hardly a household name; it is a small, relatively new, Washington-based nongovernmental organization (NGO) employing three full-time researchers. Nonetheless, its stated mission of promoting "vigorous" public debate about economic problems was accomplished with a vengeance when it came to the Multilateral Agreement on Investment (MAI).

The primary objective of the governments negotiating the MAI was to facilitate international investment by ensuring that host governments treat all foreign and domestic firms similarly. While the 29 wealthy nations comprising the Organization for Economic Cooperation and Development (OECD) were busy negotiating, however, other unexpected—and unwelcome—participants were looking over their shoulders.

The Preamble Collaborative was one of more than 600 organizations in nearly 70 countries expressing vehement opposition to the treaty, often in apocalyptic terms. The collaborative's extensive World Wide Web site—featuring fact sheets, congressional testimony, position papers, and issue briefs—was part of a tidal wave of electronically amplified public opposition to the MAI. It was cited on more than 50 other different Web sites and 200 news group postings by what Guy de .

{comprising}

STEPHEN J. KOBRIN *is director of the Lauder Institute of Management and International Studies and William Wurster professor of multinational management at the Wharton School. John Muir MacPherson helped with the research for this article.*

Jonquières of the *Financial Times* has described as "network guerrillas"—a horde of vigilantes who ambushed the negotiations.

The ambush began in February 1997 when an early draft of the treaty, replete with numerous contradictions, was leaked to Public Citizen, a Washington-based public interest group founded by Ralph Nader, and then immediately published on the Web. Up to that point, negotiations had been conducted in relative isolation—they were not reported in depth in the *New York Times*, *Washington Post*, *Los Angeles Times*, or *Christian Science Monitor*. Suddenly, what had been a working document among 29 parties became available to anyone with a computer and a modem.

And everyone with a computer and a modem got involved. OECD representatives quickly became the targets of unprecedented scrutiny. "If a negotiator says something to someone over a glass of wine, we'll have it on the Internet within an hour, all over the world," boasted the head of the Council of Canadians, a citizens' interest group claiming more than 100,000 members.

The MAI was denounced as a major and immediate threat to democracy, sovereignty, the environment, human rights, and economic development. A coalition of strange bedfellows arose in opposition to the treaty, including the AFL-CIO, Amnesty International, Australian Conservation Foundation, Friends of the Earth, Oxfam, Public Citizen, Sierra Club, Third World Network, United Steelworkers of America, Western Governors' Association, and World Development Movement. They claimed that the MAI would give corporations the "sovereign power to govern countries," make elected governments "their compliant puppets," and "radically limit our ability to promote social, economic, and environmental justice."

Three years of negotiations on the MAI screeched to a halt in late April 1998 when the OECD announced that talks would be delayed for six months. Negotiators called a time-out to allow for consultation among the parties and "with interested parts of their societies," including NGOS, business, and labor. Although some newspapers proclaimed that the Internet had sunk the MAI—a triumph of civil society over civil servants—the anti-MAI forces could not take all the credit for tabling the talks; the participants' inability to agree also played an important role. The short preamble to the treaty, for example, contains 17 footnotes expressing the concerns of one or more delegations. The latest draft contains almost 50 pages of country-specific exceptions. Harmony appears hard to come by.

Yet even the MAI's negotiators concede that the NGOs had an impact. "The growing pressure from civil society further exacerbated the differences of opinion within the OECD," observed Belgium's Foreign Trade Ministry. Indeed, the battle against the MAI was not limited to the Web. Letter-writing campaigns, petition signings, and public protests caught the attention of politicians. As one European official observed, the wave of protest elevated the question of the MAI from the "level of civil servants" to the "ministerial level."

French parliamentarians heeded the cry of intellectuals who claimed that the MAI would permit crude foreign culture to permeate their country. The government of New Zealand pledged not to sign the treaty, as anti–MAI protesters held a mock auction, selling off the country to McDonald's, Pepsi, and Shell outside the Foreign Affairs and Trade building.

Why has the MAI generated such broad, intense, and vehement opposition? What turned this arcane treaty that attempts to institutionalize liberalization of international investment flows into a source of very real concern—indeed fear—among people who normally might confuse FDI with the FBI?

The story of the MAI is a cautionary tale about the impact of an electronically networked global civil society.

On one level, the story of the MAI is a cautionary tale about the impact of an electronically networked global civil society. The days of negotiating international treaties behind closed doors are numbered, if not over. A much broader range of groups will have to be included in the globalization debate, and much more thought will have to be given to how nonparticipants will interpret international negotiations and agreements.

On a broader level, the battle over the MAI is a reminder that although the pace and structure of globalization are still open to debate, the phenomenon of globalization is a fait accompli. For the past decade, NGOs and politicians—conservative, leftist, and populist—have railed against globalization and sought to promote alternatives. But in doing so, antiglobalization activists and advocacy groups have become transnational actors themselves. Both international investors and the electronically networked opposition to the MAI are manifestations of globalization; both compromise the concept of national sovereignty and local control.

THE TREATY IN QUESTION

Over the last decade, worldwide market liberalization has prompted dramatic growth in foreign direct investment. The global stock of FDI increased fourfold from 1982 to 1994, and annual flows of cross-border investment reached an all-time high of $350 billion in 1996. Consequently, Western industrialized countries once again began to focus on the governance of international investment. By the mid-1990s, there was general agreement among these nations on the need for a global, multilateral framework to replace the roughly 1,600 bilateral investment treaties in existence.

In May 1995, the OECD began negotiations on the MAI to provide a comprehensive framework for international investment, institutionalizing liberalization while providing for the protection of investment and the resolution of disputes. The MAI was to provide the framework for international investment that the General Agreement on Tariffs and Trade (GATT) provided for international trade. It was intended as a free-standing international treaty open to all OECD members and to accession by non–OECD members.

Although the choice of the OECD as a venue was controversial, proponents argued that since its members represented 85 percent of all FDI outflows, having them negotiate among themselves would produce a better agreement. In practice, this meant shutting out the developing countries.

National treatment and most-favored-nation (MFN) obligations are at the core of the MAI. These two provisions also form the basis for much of the opposition. MFN obligations, which are borrowed from trade agreements, stipulate that benefits extended to investors from one country be extended to all. National treatment requires that foreign investors not be discriminated against by national law and regulation, and that foreign and national firms be treated similarly in like circumstances. (There are a large number of country-specific exceptions to national treatment under discussion.)

In contrast to most other investment treaties, which protect investments only after they are made, the MAI also covers the "pre-establishment" phase of investment. It applies to a country's framework of law and regulation governing foreign investment, intending both to institutionalize liberalization and protect specific investment projects.

The MAI is complex, incomplete, and very much a "work in progress by committee." It is replete with arcane language, alternative phrasings,

Rules for Corporations. . .

Major provisions of the Multilateral Agreement on Investment (MAI):

- A broad definition of investment to include investment in stocks and bonds, as well as foreign direct investment and contract rights, intellectual property, real estate, and "claims to money."
- Very strict limits on "performance requirements"—laws governing such matters as the obligation to have a certain level of local content, exports, local hiring, local research and development, transfer of technology, and domestic equity participation, among others. This provision is still under negotiation and exceptions to protect the environment and secure compliance with local law are also being considered.
- Limits on expropriation subject to the "usual" justifications and conditions: a public purpose; nondiscriminatory application; due process; and prompt, adequate, and effective compensation. The phrasing, however, is quite broad, including both nationalization and "any. . .measures having equivalent effect."
- Free transfer or repatriation of capital, profits, interest payments, expropriation settlements, and the like.
- Dispute settlement provisions that establish an international tribunal to arbitrate between countries and give private investors the standing to sue a country in its courts for breach of the agreement or to bring action in an international tribunal.
- Provisions that require countries to "roll back" existing laws or regulations that are not in accordance with the MAI and refrain from passing new laws that contradict it.
- Specific application of nondiscrimination or national treatment to privatization, monopoly regulation, and access to minerals and raw materials.

—S.J.K.

exceptions, paragraphs that may or may not be included in the final draft, and objections on the part of one or more delegations.

That being said, a reasonably thorough reading of the MAI draft raises some concerns. The treaty appears somewhat one-sided, focusing—as critics note—on the rights of investors and not on their obligations. The provisions banning performance requirements—laws that require investors to meet certain conditions if they want to establish an enterprise in a specific location—appear much too comprehensive.

Likewise, the discussion of expropriation seems much too loosely worded: It could be interpreted to cover any act of government that affects operations or profit.

Although others will find similar concerns in other parts of the treaty, many observers will probably wonder if three years of negotiations—and the subsequent addition of footnotes, amendments, exceptions, and objections—have diluted it to the point where it is too little, too late. In any event, the MAI does not appear to be the stuff of which revolutions are made. Yet, it has sparked a widespread grassroots opposition taking the form of Web sites, news groups, bumper stickers, newspaper advertisements ("Should Corporations Govern the World?"), and even street demonstrations. It prompted 14,000 people to write the U.S. State Department. What is going on here?

NAFTA ON STEROIDS?

Much of the public concern focuses on five provisions of the MAI: treating foreign corporations as national firms; extending benefits given to foreign investors from any country to all (MFN); the ban on performance requirements; the expropriation clause; and the right of investors to sue governments. The expropriation clause, for example, bars both direct nationalization of assets and "any other measure or measures having equivalent effect. . . ." The clause has been widely interpreted as barring any law or regulation that impedes, or will impede, an investor's right to make a profit. Thus, opponents argue that environmental, health, or workers' rights legislation that could threaten profits would be interpreted as "expropriation" and prohibited by the treaty. The Sierra Club, for instance, argues that the MAI might prohibit bans on exports of raw (unprocessed) logs from some national forests.

MAI opponents believe that the expropriation clause, in conjunction with the "unprecedented legal standing" accorded private investors under the dispute resolution provisions of the MAI, will allow multinational firms to sue any government that takes any action whatsoever that might impede their right to make a profit. The U.S.-based Ethyl Corporation's suit against the Canadian government is widely cited as a warning of things to come: When the Canadian parliament banned a fuel additive produced by Ethyl for environmental and health reasons, the company sued for damages, claiming that Canada violated its North American Free Trade Agreement (NAFTA) commitments on expropriation and compensation, performance requirements, and national treatment.

Another widespread concern is that the MFN clause would prohibit boycotts against countries that violate human rights or the environment. The assumption, which is hard to square with 50 years of experience with the GATT, is that since MFN requires treating all investing countries alike, it would bar "discrimination" against any of them. In other words, if the MAI had been in force, apartheid would still be with us, Nelson Mandela would still be in jail, and it would be impossible to single out future South Africas for sanctions.

Moreover, although the central purpose of the treaty is to level the playing field among nations, the issue of equality looms large on the anti–MAI agenda. Developing nations, already seething over their exclusion from negotiations, are adamant that their governments retain the right to regulate FDI. Martin Khor, director of the Malaysia-based Third World Network, contends that the economies of most developing countries were "shaped to the advantage of foreign companies and financial institutions." Consequently, local firms might require special treatment for a long time "before they can compete on more balanced terms" with large foreign companies.

Opponents assume that the MAI will markedly limit national and local governments' abilities to regulate and legislate protection of the environment, health, natural resources, and local firms and citizens. That at a minimum, local firms, investors, workers, and citizens cannot be favored over or protected from foreign investors, even when it is warranted. While some of these concerns do arise logically from the treaty, there are a large number of exceptions being negotiated to deal with specific issues.

Many of the opponents' arguments stretch concepts such as national treatment and MFN to the breaking point. Anti–MAI activists worry that the essence of the democratic process will be violated and any action that interferes with the profits of foreign investors will be taboo. These fears clearly transcend concerns about the impact of specific clauses in the MAI treaty.

In fact, they reflect the reality of globalization. An electronically integrated global economy may not obliterate the nation-state, but it will affect how states are structured and how they function. As many academics have argued, power is shifting from states to the market. There has been a loss of local control over economies and economic actors and the preservation of democratic processes has become a real concern.

. . . or Corporate Rule?

"The MAI [Multilateral Agreement on Investment], if ratified, will serve as a Charter of Rights and Freedoms for transnational corporations against citizens and the earth, and represents a grave threat to democracy in Canada and around the world."

– *MAI: The Multilateral Agreement on Investment and the Threat to Canadian Sovereignty*, by Maude Barlow and Tony Clarke

"[The MAI is] one of the greatest threats ever to the economic development and national sovereignty of countries of the South."

– Dr. Chandra Muzaffar, director, Just World Trust

"If the OECD gets its way, the British government will never again be permitted to restrain the rapacity of the private sector."

– Environmental advocate George Monbiot, letter to the London *Guardian*, April 15, 1997

"I'm scared. And no, I'm not scared of the simple, everyday things that Grade 12 students normally fear. . . . What I fear can be expressed in four words. . . Multinational Agreement on Investment."

– High-school student Alan Slipp, from a speech given at the Annual Lion's Club Speakout in Nova Scotia

"Frances [sic] Fukuyama may be satisfied that the current winning streak of market ideology heralds the 'end of history.' The corporations, however, want to put it in writing."

– Scott Nova and Michelle Sforza-Roderick, Preamble Center for Public Policy

"Under MAI, local, regional, and federal governments could no longer make low-interest loans to local businesses, cut taxes for businesses that hire members of local communities, or give minority-owned or environmentally conscious companies preference in the awarding of public-works contracts."

– Gabriel Roth, *San Francisco Bay Guardian*, October 15, 1997

"The MAI takes us so much further down [the] road [of corporate dominance] that we might never conceivably return until we are driven to a social and political revolution at a global scale."

– Janice Harvey, *Telegraph-Journal*, New Brunswick, Canada, April 30, 1997

In the midst of such traumatic and systemic change, the MAI serves as a lightning rod. Sometimes described as "NAFTA on steroids," it has become a visible rallying point for opposition to a global economy.

Although much of this opposition is polemic, it would be a mistake to dismiss it, as a State Department spokesman did when he derided the anti–MAI activists as the "black helicopter crowd." A number of the concerns about the MAI are tangible manifestations of a more general anxiety about globalization:

- **Economic globalization has gone too far.** The level of liberalization embodied by the MAI is seen as inappropriate, especially for many developing countries. Completely free flows of capital may not be desirable everywhere.
- **The state-market balance of power has shifted and corporations have too much power.** The MAI ~~allegedly~~ seems to put~~s~~ business interests above all competing social concerns. Corporations have new rights under the MAI but no corresponding obligations to workers, consumers, or the environment. The MAI will put multinational corporations beyond the reach of national and local economic governance.
- **Globalization compromises national sovereignty.** The MAI is said to be part of a transnational regulatory framework that will override national jurisdiction. This concern was given form by the now-infamous claim of World Trade Organization director general Renato Ruggiero that, "We are writing the constitution of a single global economy." Not everyone thinks this is a great idea; many see the MAI as "a constitution for the largest corporations to rule the world," with elected governments acting "as their compliant puppets."
- **Globalization reduces transparency and accountability, shifting power from elected national (and local) officials to nonelected (trade) bureaucrats and international officials.** There is an assumption here that OECD secrecy was both purposeful and necessary. The MAI, in the words of one critic, is "like a political Dracula, [which] simply cannot survive sunlight."
- **Globalization limits national and local economic policy choices.** The MAI is seen as strictly limiting national regulation of national economies. It "was developed to enable investors in multinational corporations to discourage any legislation issued by national or even subnational governments that foreign investors perceive as against their profit objectives."

ADVERSARIES OR COLLABORATORS?

There are important lessons to be learned on both sides of this issue by national governments and international organizations concerned with economic governance in a global age, and by organizations and activists concerned with globalization's impact on individuals, communities, and the environment.

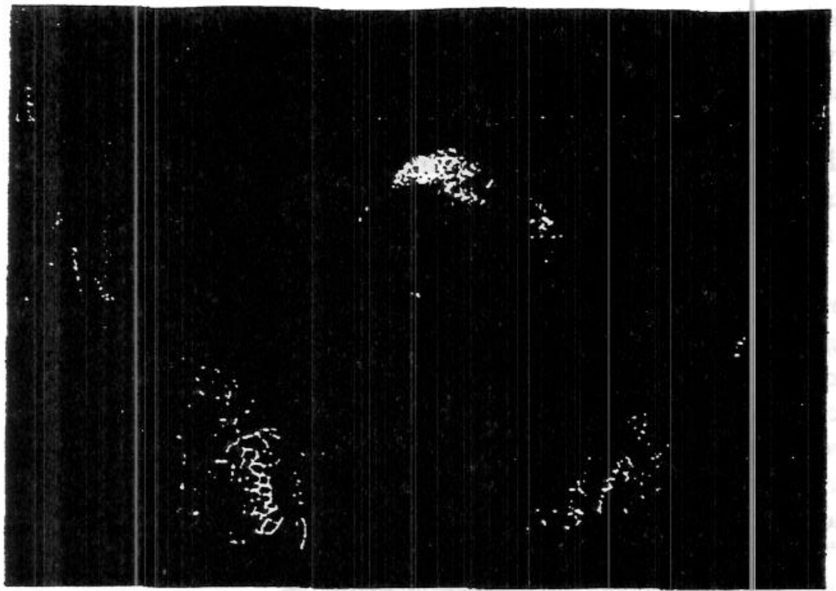
For starters, the opponents of the MAI and, more broadly speaking, of globalization, cannot stem the tide by yelling at the surf, by wishing for a counterfactual world where globalization does not exist. They cannot pick and choose, selecting the electronic global village—the emergence of global civil society—as a good thing and increased economic integration or a loss of local control as a bad thing to be unambiguously opposed. It is much easier to talk about secret negotiations where plots are hatched for corporate oligarchies to rule the world than to deal with the complex problems that globalization poses for equality of opportunity, the environment, and the democratic process.

But proponents of economic globalization must also learn the same lesson. As the OECD negotiators found out the hard way, NGOs and other advocacy groups are now electronically networked across national borders. Real secrecy will be hard to achieve when information can be broadly disseminated with the push of a button. National governments and international organizations must assume that news will be leaked and published on the Web, that negotiations will take place in public.

As a corollary, globalization cannot be a top-down or élite-driven project. Policymakers cannot assume that all reasonable people share their assumptions and values. Not everyone believes that a constitution for a new global economy or a new international economic order is desirable. Not everyone believes that an open international economy, with free flows of trade, capital, and direct investment promotes the general welfare.

Consequently, there will be a continuous public referendum of sorts on these issues. Much more thought has to be given to how the phrasing of debates and agreements will be interpreted by nonparticipants. Public affairs have become critical. The secretary general of the OECD now concedes that a "strategy on information, communication and explication" is necessary.

The reason that opposition to the MAI has been so successful is that the treaty has been presented on the Internet in terms that are imme-



If the MAI had existed, would he be in prison right now?

diate, meaningful, and threatening to a very large number of disparate individuals and groups. Download the report prepared by the Western Governors' Association, for example, and you will read a detailed, state-by-state listing of specific laws and ordinances that might be threatened if the United States were subject to the MAI.

Much of the anti–MAI sentiment on the Internet presents barely credible worst-case scenarios as fact. As the OECD discovered, much to its chagrin, there are no controls on the Net over who can "publish" or what they can say. Although some of the arguments—the Preamble Collaborative's, for example—are balanced and reasoned, most of the rest are neither. The MAI deals with difficult and often technical issues, and considerable disagreement remains among the parties to the treaty. The Internet is a medium where the most extreme statements attract attention; where an argument scrolling down a computer screen may garner authority it does not deserve.

The interwoven nature of the Web—the links and hypertext—also provides an effective dissemination process. Whatever strikes a chord gets picked up and repeated. The Internet allows anti–MAI activists to reach large numbers of people, at little or no cost, who normally would never hear of an investment treaty negotiated in a far away place and would never think that it might affect them directly.

When the OECD realized that anyone looking for the MAI on the Web encountered only hostile sites, it was forced to establish an "official" MAI Web site. But current drafts of the treaty, commentaries, and memoranda were no match for apocalyptic claims of environmental ruin and corporate rule. Not everyone is born with an innate understanding of trade theory or macroeconomics. An open international economy entails costs as well as benefits, and both need to be carefully explained. Generating public support for treaties such as the MAI will be difficult and costly. However, the alternative is now clear.

The Information Age gives new powers, and new responsibilities, to the wide variety of actors forming the core of the new global, electronically interconnected civil society. It is a large virtual community that unites like-minded groups across great distances; some estimates have put the total number of transnational NGOs at 20,000. [See P.J. Simmons' article on page TK.] As one observer of the MAI debate has noted, the NGOS have "tasted blood" and will be back. No longer satisfied with simply opposing whatever proposals the negotiators happen to place on the table, there is growing talk among them that their organizations should play a direct role in drafting the agenda.

But will the process remain adversarial or will it become more collaborative? These two extremes are not viable. It will be increasingly difficult to conduct international negotiations in private, much less in secret, or to impose globalization as an élite-driven project. Yet conducting an "electronic" public referendum on every issue simply will not work. In a global economy, all politics cannot be local. A middle ground must emerge that allows for both broader public involvement and some semblance of efficient and effective global governance. An electronically integrated global civil society and a global economy are two sides of the same coin.

WANT TO KNOW MORE?

For a thoughtful discussion of the impact of an electronic, global civil society on political authority and power, see David Rothkopf's **"Cyberpolitik: The Changing Nature of Power In the Information Age"** (*Journal of International Affairs*, Spring 1998) and Jessica Mathews' **"Power Shift"** (*Foreign Affairs*, January/February 1997). Manuel Castells' three-volume work, ***The Information Age: Economy, Society*** *and Culture* (Oxford: Blackwell Publishers) provides a comprehensive and extended discussion of the impact of the information revolution on our societies, politics, and economies. The last volume, ***End of Millennium*** (1998), is particularly relevant.

The best single source of information on MAI and all of the arguments for and against it, is the Web itself. The "official" MAI site of the OECD provides the current draft of the full negotiating text, commentaries, reports, seminars, publications, articles, speeches, and much more. The MAI Not! site of the Ontario Public Interest Research Group is a focal point for anti–MAI materials. The site contains links to a vast collection of MAI materials and links to many of the other active anti–MAI groups. Perhaps the most reasoned, balanced, and comprehensive arguments against MAI are those of the Preamble Collaborative, whose Web site features numerous news briefs, fact-sheets, and extensive analyses of the agreement. Tony Clarke's paper **"MAI–DAY! The Corporate Rule Treaty"** can be found on the Canadian Centre for Policy Alternatives' Web site. A number of the major environmental organizations including Friends of the Earth and the Sierra Club have a good deal of anti–MAI material on their sites.

For links to these and other Web sites, as well as a comprehensive index of related articles, access **www.foreignpolicy.com**.

LIPPINCOTT LIBRARY
DOCUMENT DELIVERY SERVICE

WHARTON REPROGRAPHICS

2.

The Clash of Civilizations?

Samuel P. Huntington

THE NEXT PATTERN OF CONFLICT

WORLD POLITICS IS entering a new phase, and intellectuals have not hesitated to proliferate visions of what it will be—the end of history, the return of traditional rivalries between nation states, and the decline of the nation state from the conflicting pulls of tribalism and globalism, among others. Each of these visions catches aspects of the emerging reality. Yet they all miss a crucial, indeed a central, aspect of what global politics is likely to be in the coming years.

It is my hypothesis that the fundamental source of conflict in this new world will not be primarily ideological or primarily economic. The great divisions among humankind and the dominating source of conflict will be cultural. Nation states will remain the most powerful actors in world affairs, but the principal conflicts of global politics will occur between nations and groups of different civilizations. The clash of civilizations will dominate global politics. The fault lines between civilizations will be the battle lines of the future.

Conflict between civilizations will be the latest phase in the evolution of conflict in the modern world. For a century and a half after the emergence of the modern international system with the Peace of Westphalia, the conflicts of the Western world were largely among

---

SAMUEL P. HUNTINGTON is the Eaton Professor of the Science of Government and Director of the John M. Olin Institute for Strategic Studies at Harvard University. This article is the product of the Olin Institute's project on "The Changing Security Environment and American National Interests."

princes—emperors, absolute monarchs and constitutional monarchs attempting to expand their bureaucracies, their armies, their mercantilist economic strength and, most important, the territory they ruled. In the process they created nation states, and beginning with the French Revolution the principal lines of conflict were between nations rather than princes. In 1793, as R. R. Palmer put it, "The wars of kings were over; the wars of peoples had begun." This nineteenth-century pattern lasted until the end of World War I. Then, as a result of the Russian Revolution and the reaction against it, the conflict of nations yielded to the conflict of ideologies, first among communism, fascism-Nazism and liberal democracy, and then between communism and liberal democracy. During the Cold War, this latter conflict became embodied in the struggle between the two superpowers, neither of which was a nation state in the classical European sense and each of which defined its identity in terms of its ideology.

These conflicts between princes, nation states and ideologies were primarily conflicts within Western civilization, "Western civil wars," as William Lind has labeled them. This was as true of the Cold War as it was of the world wars and the earlier wars of the seventeenth, eighteenth and nineteenth centuries. With the end of the Cold War, international politics moves out of its Western phase, and its centerpiece becomes the interaction between the West and non-Western civilizations and among non-Western civilizations. In the politics of civilizations, the peoples and governments of non-Western civilizations no longer remain the objects of history as targets of Western colonialism but join the West as movers and shapers of history.

THE NATURE OF CIVILIZATIONS

DURING THE COLD WAR the world was divided into the First, Second and Third Worlds. Those divisions are no longer relevant. It is far more meaningful now to group countries not in terms of their political or economic systems or in terms of their level of economic development but rather in terms of their culture and civilization.

What do we mean when we talk of a civilization? A civilization is a cultural entity. Villages, regions, ethnic groups, nationalities, religious groups, all have distinct cultures at different levels of cultural heterogeneity. The culture of a village in southern Italy may be different from that of a village in northern Italy, but both will share in a common Italian culture that distinguishes them from German villages. European communities, in turn, will share cultural features that distinguish them from Arab or Chinese communities. Arabs, Chinese and Westerners, however, are not part of any broader cultural entity. They constitute civilizations. A civilization is thus the highest cultural grouping of people and the broadest level of cultural identity people have short of that which distinguishes humans from other species. It is defined both by common objective elements, such as language, history, religion, customs, institutions, and by the subjective self-identification of people. People have levels of identity: a resident of Rome may define himself with varying degrees of intensity as a Roman, an Italian, a Catholic, a Christian, a European, a Westerner. The civilization to which he belongs is the broadest level of identification with which he intensely identifies. People can and do redefine their identities and, as a result, the composition and boundaries of civilizations change.

Civilizations may involve a large number of people, as with China ("a civilization pretending to be a state," as Lucian Pye put it), or a very small number of people, such as the Anglophone Caribbean. A civilization may include several nation states, as is the case with Western, Latin American and Arab civilizations, or only one, as is the case with Japanese civilization. Civilizations obviously blend and overlap, and may include subcivilizations. Western civilization has two major variants, European and North American, and Islam has its Arab, Turkic and Malay subdivisions. Civilizations are nonetheless meaningful entities, and while the lines between them are seldom sharp, they are real. Civilizations are dynamic; they rise and fall; they divide and merge. And, as any student of history knows, civilizations disappear and are buried in the sands of time.

Westerners tend to think of nation states as the principal actors in global affairs. They have been that, however, for only a few centuries. The broader reaches of human history have been the history of civilizations. In *A Study of History*, Arnold Toynbee identified 21 major civilizations; only six of them exist in the contemporary world.

WHY CIVILIZATIONS WILL CLASH

CIVILIZATION IDENTITY will be increasingly important in the future, and the world will be shaped in large measure by the interactions among seven or eight major civilizations. These include Western, Confucian, Japanese, Islamic, Hindu, Slavic-Orthodox, Latin American and possibly African civilization. The most important conflicts of the future will occur along the cultural fault lines separating these civilizations from one another.

Why will this be the case?

First, differences among civilizations are not only real; they are basic. Civilizations are differentiated from each other by history, language, culture, tradition and, most important, religion. The people of different civilizations have different views on the relations between God and man, the individual and the group, the citizen and the state, parents and children, husband and wife, as well as differing views of the relative importance of rights and responsibilities, liberty and authority, equality and hierarchy. These differences are the product of centuries. They will not soon disappear. They are far more fundamental than differences among political ideologies and political regimes. Differences do not necessarily mean conflict, and conflict does not necessarily mean violence. Over the centuries, however, differences among civilizations have generated the most prolonged and the most violent conflicts.

> The conflicts of the future will occur along the cultural fault lines separating civilizations.

Second, the world is becoming a smaller place. The interactions between peoples of different civilizations are increasing; these increasing interactions intensify civilization consciousness and awareness of differences between civilizations and commonalities within civilizations. North African immigration to France generates hostility among Frenchmen and at the same time increased receptivity to immigration by "good" European Catholic Poles. Americans react far more negatively to Japanese investment than to larger investments from Canada and European countries. Similarly, as Donald Horowitz has pointed out, "An Ibo may be ... an Owerri Ibo or an Onitsha Ibo in what was the Eastern region of Nigeria. In Lagos, he is simply an Ibo. In London, he is a Nigerian. In New York, he is an African." The interactions among peoples of different civilizations enhance the civilization-consciousness of people that, in turn, invigorates differences and animosities stretching or thought to stretch back deep into history.

Third, the processes of economic modernization and social change throughout the world are separating people from longstanding local identities. They also weaken the nation state as a source of identity. In much of the world religion has moved in to fill this gap, often in the form of movements that are labeled "fundamentalist." Such movements are found in Western Christianity, Judaism, Buddhism and Hinduism, as well as in Islam. In most countries and most religions the people active in fundamentalist movements are young, college-educated, middle-class technicians, professionals and business persons. The "unsecularization of the world," George Weigel has remarked, "is one of the dominant social facts of life in the late twentieth century." The revival of religion, "la revanche de Dieu," as Gilles Kepel labeled it, provides a basis for identity and commitment that transcends national boundaries and unites civilizations.

Fourth, the growth of civilization-consciousness is enhanced by the dual role of the West. On the one hand, the West is at a peak of power. At the same time, however, and perhaps as a result, a return to the roots phenomenon is occurring among non-Western civilizations. Increasingly one hears references to trends toward a turning inward and "Asianization" in Japan, the end of the Nehru legacy and the "Hinduization" of India, the failure of Western ideas of socialism and nationalism and hence "re-Islamization" of the Middle East, and now a debate over Westernization versus Russianization in Boris Yeltsin's country. A West at the peak of its power confronts non-Wests that increasingly have the desire, the will and the resources to shape the world in non-Western ways.

In the past, the elites of non-Western societies were usually the people who were most involved with the West, had been educated at Oxford, the Sorbonne or Sandhurst, and had absorbed Western attitudes and values. At the same time, the populace in non-Western countries often remained deeply imbued with the indigenous culture. Now, however, these relationships are being reversed. A de-Westernization and indigenization of elites is occurring in many non-Western countries at the same time that Western, usually American, cultures, styles and habits become more popular among the mass of the people.

Fifth, cultural characteristics and differences are less mutable and hence less easily compromised and resolved than political and economic ones. In the former Soviet Union, communists can become democrats, the rich can become poor and the poor rich, but Russians cannot become Estonians and Azeris cannot become Armenians. In class and ideological conflicts, the key question was "Which side are you on?" and people could and did choose sides and change sides. In conflicts between civilizations, the question is "What are you?" That is a given that cannot be changed. And as we know, from Bosnia to the Caucasus to the Sudan, the wrong answer to that question can mean a bullet in the head. Even more than ethnicity, religion discriminates sharply and exclusively among people. A person can be half-French and half-Arab and simultaneously even a citizen of two countries. It is more difficult to be half-Catholic and half-Muslim.

Finally, economic regionalism is increasing. The proportions of total trade that were intraregional rose between 1980 and 1989 from 51 percent to 59 percent in Europe, 33 percent to 37 percent in East Asia, and 32 percent to 36 percent in North America. The importance of regional economic blocs is likely to continue to increase in the future. On the one hand, successful economic regionalism will reinforce civilization-consciousness. On the other hand, economic regionalism may succeed only when it is rooted in a common civilization. The European Community rests on the shared foundation of European culture and Western Christianity. The success of the North American Free Trade Area depends on the convergence now underway of Mexican, Canadian and American cultures. Japan, in contrast, faces difficulties in creating a comparable economic entity in East Asia because Japan is a society and civilization unique to itself. However strong the trade and investment links Japan may develop with other East Asian countries, its cultural differences with those countries inhibit and perhaps preclude its promoting regional economic integration like that in Europe and North America.

Common culture, in contrast, is clearly facilitating the rapid expansion of the economic relations between the People's Republic of China and Hong Kong, Taiwan, Singapore and the overseas Chinese communities in other Asian countries. With the Cold War over, cultural commonalities increasingly overcome ideological differences, and mainland China and Taiwan move closer together. If cultural commonality is a prerequisite for economic integration, the principal East Asian economic bloc of the future is likely to be centered on China. This bloc is, in fact, already coming into existence. As Murray Weidenbaum has observed,

> Despite the current Japanese dominance of the region, the Chinese-based economy of Asia is rapidly emerging as a new epicenter for industry, commerce and finance. This strategic area contains substantial amounts of technology and manufacturing capability (Taiwan), outstanding entrepreneurial, marketing and services acumen (Hong Kong), a fine communications network (Singapore), a tremendous pool of financial capital (all three), and very large endowments of land, resources and labor (mainland China).... From Guangzhou to Singapore, from Kuala Lumpur to Manila, this influential network—often based on extensions of the traditional clans—has been described as the backbone of the East Asian economy.[1]

Culture and religion also form the basis of the Economic Cooperation Organization, which brings together ten non-Arab Muslim countries: Iran, Pakistan, Turkey, Azerbaijan, Kazakhstan, Kyrgyzstan, Turkmenistan, Tadjikistan, Uzbekistan and Afghanistan. One impetus to the revival and expansion of this organization, founded originally in the 1960s by Turkey, Pakistan and Iran, is the realization by the leaders of several of these countries that they had no chance of admission to the European Community. Similarly, Caricom, the Central American Common Market and Mercosur rest on common cultural foundations. Efforts to build a broader Caribbean-Central American economic entity bridging the Anglo-Latin divide, however, have to date failed.

---

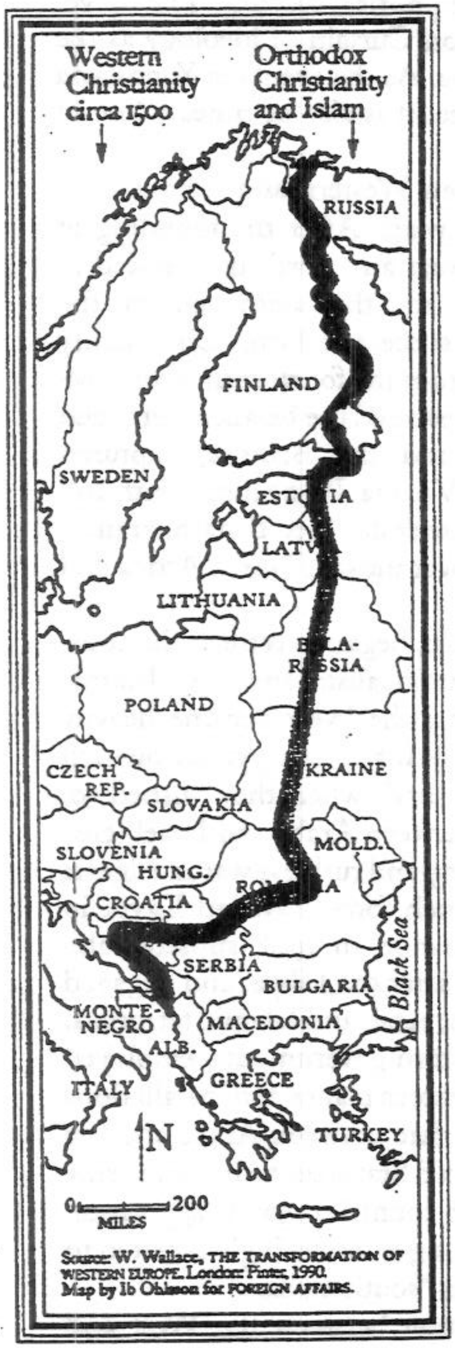
[1] Murray Weidenbaum, *Greater China: The Next Economic Superpower?*, St. Louis: Washington University Center for the Study of American Business, Contemporary Issues, Series 57, February 1993, pp. 2-3.

As people define their identity in ethnic and religious terms, they are likely to see an "us" versus "them" relation existing between themselves and people of different ethnicity or religion. The end of ideologically defined states in Eastern Europe and the former Soviet Union permits traditional ethnic identities and animosities to come to the fore. Differences in culture and religion create differences over policy issues, ranging from human rights to immigration to trade and commerce to the environment. Geographical propinquity gives rise to conflicting territorial claims from Bosnia to Mindanao. Most important, the efforts of the West to promote its values of democracy and liberalism as universal values, to maintain its military predominance and to advance its economic interests engender countering responses from other civilizations. Decreasingly able to mobilize support and form coalitions on the basis of ideology, governments and groups will increasingly attempt to mobilize support by appealing to common religion and civilization identity.

The clash of civilizations thus occurs at two levels. At the micro-level, adjacent groups along the fault lines between civilizations struggle, often violently, over the control of territory and each other. At the macro-level, states from different civilizations compete for relative military and economic power, struggle over the control of international institutions and third parties, and competitively promote their particular political and religious values.

THE FAULT LINES BETWEEN CIVILIZATIONS

THE FAULT LINES between civilizations are replacing the political and ideological boundaries of the Cold War as the flash points for crisis and bloodshed. The Cold War began when the Iron Curtain divided Europe politically and ideologically. The Cold War ended with the end of the Iron Curtain. As the ideological division of Europe has disappeared, the cultural division of Europe between Western Christianity, on the one hand, and Orthodox Christianity


Western Christianity circa 1500
Orthodox Christianity and Islam
RUSSIA
FINLAND
SWEDEN
ESTONIA
LATVIA
LITHUANIA
BELARUSSIA
POLAND
CZECH REP.
SLOVAKIA
UKRAINE
MOLD.
SLOVENIA
HUNG.
CROATIA
ROMANIA
SERBIA
BULGARIA
Black Sea
MONTENEGRO
MACEDONIA
ALB.
GREECE
ITALY
TURKEY
N
0 200
MILES

Source: W. Wallace, THE TRANSFORMATION OF WESTERN EUROPE. London: Pinter, 1990.
Map by Ib Ohlsson for FOREIGN AFFAIRS.

and Islam, on the other, has reemerged. The most significant dividing line in Europe, as William Wallace has suggested, may well be the eastern boundary of Western Christianity in the year 1500. This line runs along what are now the boundaries between Finland and Russia and between the Baltic states and Russia, cuts through Belarus and Ukraine separating the more Catholic western Ukraine from Orthodox eastern Ukraine, swings westward separating Transylvania from the rest of Romania, and then goes through Yugoslavia almost exactly along the line now separating Croatia and Slovenia from the rest of Yugoslavia. In the Balkans this line, of course, coincides with the historic boundary between the Hapsburg and Ottoman empires. The peoples to the north and west of this line are Protestant or Catholic; they shared the common experiences of European history—feudalism, the Renaissance, the Reformation, the Enlightenment, the French Revolution, the Industrial Revolution; they are generally economically better off than the peoples to the east; and they may now look forward to increasing involvement in a common European economy and to the consolidation of democratic political systems. The peoples to the east and south of this line are Orthodox or Muslim; they historically belonged to the Ottoman or Tsarist empires and were only lightly touched by the shaping events in the rest of Europe; they are generally less advanced economically; they seem much less likely to develop stable democratic political systems. The Velvet Curtain of culture has replaced the Iron Curtain of ideology as the most significant dividing line in Europe. As the events in Yugoslavia show, it is not only a line of difference; it is also at times a line of bloody conflict.

Conflict along the fault line between Western and Islamic civilizations has been going on for 1,300 years. After the founding of Islam, the Arab and Moorish surge west and north only ended at Tours in 732. From the eleventh to the thirteenth century the Crusaders attempted with temporary success to bring Christianity and Christian rule to the Holy Land. From the fourteenth to the seventeenth century, the Ottoman Turks reversed the balance, extended their sway over the Middle East and the Balkans, captured Constantinople, and twice laid siege to Vienna. In the nineteenth and early twentieth centuries as Ottoman power declined Britain, France, and Italy established Western control over most of North Africa and the Middle East.

After World War II, the West, in turn, began to retreat; the colonial empires disappeared; first Arab nationalism and then Islamic fundamentalism manifested themselves; the West became heavily dependent on the Persian Gulf countries for its energy; the oil-rich Muslim countries became money-rich and, when they wished to, weapons-rich. Several wars occurred between Arabs and Israel (created by the West). France fought a bloody and ruthless war in Algeria for most of the 1950s; British and French forces invaded Egypt in 1956; American forces went into Lebanon in 1958; subsequently American forces returned to Lebanon, attacked Libya, and engaged in various military encounters with Iran; Arab and Islamic terrorists, supported by at least three Middle Eastern governments, employed the weapon of the weak and bombed Western planes and installations and seized Western hostages. This warfare between Arabs and the West culminated in 1990, when the United States sent a massive army to the Persian Gulf to defend some Arab countries against aggression by another. In its aftermath NATO planning is increasingly directed to potential threats and instability along its "southern tier."

This centuries-old military interaction between the West and

Islam is unlikely to decline. It could become more virulent. The Gulf War left some Arabs feeling proud that Saddam Hussein had attacked Israel and stood up to the West. It also left many feeling humiliated and resentful of the West's military presence in the Persian Gulf, the West's overwhelming military dominance, and their apparent inability to shape their own destiny. Many Arab countries, in addition to the oil exporters, are reaching levels of economic and social development where autocratic forms of government become inappropriate and efforts to introduce democracy become stronger. Some openings in Arab political systems have already occurred. The principal beneficiaries of these openings have been Islamist movements. In the Arab world, in short, Western democracy strengthens anti-Western political forces. This may be a passing phenomenon, but it surely complicates relations between Islamic countries and the West.

Those relations are also complicated by demography. The spectacular population growth in Arab countries, particularly in North Africa, has led to increased migration to Western Europe. The movement within Western Europe toward minimizing internal boundaries has sharpened political sensitivities with respect to this development. In Italy, France and Germany, racism is increasingly open, and political reactions and violence against Arab and Turkish migrants have become more intense and more widespread since 1990.

On both sides the interaction between Islam and the West is seen as a clash of civilizations. The West's "next confrontation," observes M. J. Akbar, an Indian Muslim author, "is definitely going to come from the Muslim world. It is in the sweep of the Islamic nations from the Maghreb to Pakistan that the struggle for a new world order will begin." Bernard Lewis comes to a similar conclusion:

> We are facing a mood and a movement far transcending the level of issues and policies and the governments that pursue them. This is no less than a clash of civilizations—the perhaps irrational but surely historic reaction of an ancient rival against our Judeo-Christian heritage, our secular present, and the worldwide expansion of both.[2]

---

[2] Bernard Lewis, "The Roots of Muslim Rage," *The Atlantic Monthly*, vol. 266, September 1990, p. 60; *Time*, June 15, 1992, pp. 24–28.

Historically, the other great antagonistic interaction of Arab Islamic civilization has been with the pagan, animist, and now increasingly Christian black peoples to the south. In the past, this antagonism was epitomized in the image of Arab slave dealers and black slaves. It has been reflected in the on-going civil war in the Sudan between Arabs and blacks, the fighting in Chad between Libyan-supported insurgents and the government, the tensions between Orthodox Christians and Muslims in the Horn of Africa, and the political conflicts, recurring riots and communal violence between Muslims and Christians in Nigeria. The modernization of Africa and the spread of Christianity are likely to enhance the probability of violence along this fault line. Symptomatic of the intensification of this conflict was the Pope John Paul II's speech in Khartoum in February 1993 attacking the actions of the Sudan's Islamist government against the Christian minority there.

On the northern border of Islam, conflict has increasingly erupted between Orthodox and Muslim peoples, including the carnage of Bosnia and Sarajevo, the simmering violence between Serb and Albanian, the tenuous relations between Bulgarians and their Turkish minority, the violence between Ossetians and Ingush, the unremitting slaughter of each other by Armenians and Azeris, the tense relations between Russians and Muslims in Central Asia, and the deployment of Russian troops to protect Russian interests in the Caucasus and Central Asia. Religion reinforces the revival of ethnic identities and restimulates Russian fears about the security of their southern borders. This concern is well captured by Archie Roosevelt:

> Much of Russian history concerns the struggle between the Slavs and the Turkic peoples on their borders, which dates back to the foundation of the Russian state more than a thousand years ago. In the Slavs' millennium-long confrontation with their eastern neighbors lies the key to an understanding not only of Russian history, but Russian character. To understand Russian realities today one has to have a concept of the great Turkic ethnic group that has preoccupied Russians through the centuries.[3]

The conflict of civilizations is deeply rooted elsewhere in Asia. The historic clash between Muslim and Hindu in the subcontinent manifests itself now not only in the rivalry between Pakistan and India but also in intensifying religious strife within India between increasingly militant Hindu groups and India's substantial Muslim minority. The destruction of the Ayodhya mosque in December 1992 brought to the fore the issue of whether India will remain a secular democratic state or become a Hindu one. In East Asia, China has outstanding territorial disputes with most of its neighbors. It has pursued a ruthless policy toward the Buddhist people of Tibet, and it is pursuing an increasingly ruthless policy toward its Turkic-Muslim minority. With the Cold War over, the underlying differences between China and the United States have reasserted themselves in areas such as human rights, trade and weapons proliferation. These differences are unlikely to moderate. A "new cold war," Deng Xaioping reportedly asserted in 1991, is under way between China and America.

---

[3] Archie Roosevelt, *For Lust of Knowing*, Boston: Little, Brown, 1988, pp. 332-333.

The crescent-shaped Islamic bloc, from the bulge of Africa to central Asia, has bloody borders.

The same phrase has been applied to the increasingly difficult relations between Japan and the United States. Here cultural difference exacerbates economic conflict. People on each side allege racism on the other, but at least on the American side the antipathies are not racial but cultural. The basic values, attitudes, behavioral patterns of the two societies could hardly be more different. The economic issues between the United States and Europe are no less serious than those between the United States and Japan, but they do not have the same political salience and emotional intensity because the differences between American culture and European culture are so much less than those between American civilization and Japanese civilization.

The interactions between civilizations vary greatly in the extent to which they are likely to be characterized by violence. Economic competition clearly predominates between the American and European subcivilizations of the West and between both of them and Japan. On the Eurasian continent, however, the proliferation of ethnic conflict, epitomized at the extreme in "ethnic cleansing," has not been totally random. It has been most frequent and most violent between groups belonging to different civilizations. In Eurasia the great historic fault lines between civilizations are once more aflame. This is particularly true along the boundaries of the crescent-shaped Islamic bloc of nations from the bulge of Africa to central Asia. Violence also occurs between Muslims, on the one hand, and Orthodox Serbs in the Balkans, Jews in Israel, Hindus in India, Buddhists in Burma and Catholics in the Philippines. Islam has bloody borders.

CIVILIZATION RALLYING: THE KIN-COUNTRY SYNDROME

GROUPS OR STATES belonging to one civilization that become involved in war with people from a different civilization naturally try to rally support from other members of their own civilization. As the post-Cold War world evolves, civilization commonality, what H. D. S. Greenway has termed the "kin-country" syndrome, is replacing political ideology and traditional balance of power considerations as the principal basis for cooperation and coalitions. It can be seen gradually emerging in the post-Cold War conflicts in the Persian Gulf, the Caucasus and Bosnia. None of these was a full-scale war between civilizations, but each involved some elements of civilizational rallying, which seemed to become more important as the conflict continued and which may provide a foretaste of the future.

First, in the Gulf War one Arab state invaded another and then fought a coalition of Arab, Western and other states. While only a few Muslim governments overtly supported Saddam Hussein, many Arab elites privately cheered him on, and he was highly popular among large sections of the Arab publics. Islamic fundamentalist movements universally supported Iraq rather than the Western-backed governments of Kuwait and Saudi Arabia. Forswearing Arab nationalism, Saddam Hussein explicitly invoked an Islamic appeal. He and his supporters attempted to define the war as a war between civilizations. "It is not the world against Iraq," as Safar Al-Hawali, dean of Islamic Studies at the Umm Al-Qura University in Mecca, put it in a widely circulated tape. "It is the West against Islam." Ignoring the rivalry between Iran and Iraq, the chief Iranian religious leader, Ayatollah Ali Khamenei, called for a holy war against the West: "The struggle against American aggression, greed, plans and policies will be counted as a jihad, and anybody who is killed on that path is a martyr." "This is a war," King Hussein of Jordan argued, "against all Arabs and all Muslims and not against Iraq alone."

The rallying of substantial sections of Arab elites and publics behind Saddam Hussein caused those Arab governments in the anti-Iraq coalition to moderate their activities and temper their public statements. Arab governments opposed or distanced themselves from subsequent Western efforts to apply pressure on Iraq, including enforcement of a no-fly zone in the summer of 1992 and the bombing of Iraq in January 1993. The Western-Soviet-Turkish-Arab anti-Iraq coalition of 1990 had by 1993 become a coalition of almost only the West and Kuwait against Iraq.

Muslims contrasted Western actions against Iraq with the West's failure to protect Bosnians against Serbs and to impose sanctions on Israel for violating U.N. resolutions. The West, they alleged, was using a double standard. A world of clashing civilizations, however, is inevitably a world of double standards: people apply one standard to their kin-countries and a different standard to others.

Second, the kin-country syndrome also appeared in conflicts in the former Soviet Union. Armenian military successes in 1992 and 1993 stimulated Turkey to become increasingly supportive of its religious, ethnic and linguistic brethren in Azerbaijan. "We have a Turkish nation feeling the same sentiments as the Azerbaijanis," said one Turkish official in 1992. "We are under pressure. Our newspapers are full of the photos of atrocities and are asking us if we are still serious about pursuing our neutral policy. Maybe we should show Armenia that there's a big Turkey in the region." President Turgut Özal agreed, remarking that Turkey should at least "scare the Armenians a little bit." Turkey, Özal threatened again in 1993, would "show its fangs." Turkish Air Force jets flew reconnaissance flights along the Armenian border; Turkey suspended food shipments and air flights to Armenia; and Turkey and Iran announced they would not accept dismemberment of Azerbaijan. In the last years of its existence, the Soviet government supported Azerbaijan because its government was dominated by former communists. With the end of the Soviet Union, however, political considerations gave way to religious ones. Russian troops fought on the side of the Armenians, and Azerbaijan accused the "Russian government of turning 180 degrees" toward support for Christian Armenia.

Third, with respect to the fighting in the former Yugoslavia, Western publics manifested sympathy and support for the Bosnian Muslims and the horrors they suffered at the hands of the Serbs. Relatively little concern was expressed, however, over Croatian attacks on Muslims and participation in the dismemberment of Bosnia-Herzegovina. In the early stages of the Yugoslav breakup, Germany, in an unusual display of diplomatic initiative and muscle, induced the other 11 members of the European Community to follow its lead in recognizing Slovenia and Croatia. As a result of the pope's determination to provide strong backing to the two Catholic countries, the Vatican extended recognition even before the Community did. The United States followed the European lead. Thus the leading actors in Western civilization rallied behind their coreligionists. Subsequently Croatia was reported to be receiving substantial quantities of arms from Central European and other Western countries. Boris Yeltsin's government, on the other hand, attempted to pursue a middle course that would be sympathetic to the Orthodox Serbs but not alienate Russia from the West. Russian conservative and nationalist groups, however, including many legislators, attacked the government for not being more forthcoming in its support for the Serbs. By early 1993 several hundred Russians apparently were serving with the Serbian forces, and reports circulated of Russian arms being supplied to Serbia.

Islamic governments and groups, on the other hand, castigated the West for not coming to the defense of the Bosnians. Iranian leaders urged Muslims from all countries to provide help to Bosnia; in violation of the U.N. arms embargo, Iran supplied weapons and men for the Bosnians; Iranian-supported Lebanese groups sent guerrillas to train and organize the Bosnian forces. In 1993 up to 4,000 Muslims from over two dozen Islamic countries were reported to be fighting in Bosnia. The governments of Saudi Arabia and other countries felt under increasing pressure from fundamentalist groups in their own societies to provide more vigorous support for the Bosnians. By the end of 1992, Saudi Arabia had reportedly supplied substantial funding for weapons and supplies for the Bosnians, which significantly increased their military capabilities vis-à-vis the Serbs.

In the 1930s the Spanish Civil War provoked intervention from countries that politically were fascist, communist and democratic. In the 1990s the Yugoslav conflict is provoking intervention from countries that are Muslim, Orthodox and Western Christian. The parallel has not gone unnoticed. "The war in Bosnia-Herzegovina has become the emotional equivalent of the fight against fascism in the Spanish Civil War," one Saudi editor observed. "Those who died there are regarded as martyrs who tried to save their fellow Muslims."

Conflicts and violence will also occur between states and groups within the same civilization. Such conflicts, however, are likely to be less intense and less likely to expand than conflicts between civilizations. Common membership in a civilization reduces the probability of violence in situations where it might otherwise occur. In 1991 and 1992 many people were alarmed by the possibility of violent conflict between Russia and Ukraine over territory, particularly Crimea, the Black Sea fleet, nuclear weapons and economic issues. If civilization is what counts, however, the likelihood of violence between Ukrainians and Russians should be low. They are two Slavic, primarily Orthodox peoples who have had close relationships with each other for centuries. As of early 1993, despite all the reasons for conflict, the leaders of the two countries were effectively negotiating and defusing the issues between the two countries. While there has been serious fighting between Muslims and Christians elsewhere in the former Soviet Union and much tension and some fighting between Western and Orthodox Christians in the Baltic states, there has been virtually no violence between Russians and Ukrainians.

Civilization rallying to date has been limited, but it has been growing, and it clearly has the potential to spread much further. As the conflicts in the Persian Gulf, the Caucasus and Bosnia continued, the positions of nations and the cleavages between them increasingly were along civilizational lines. Populist politicians, religious leaders and the media have found it a potent means of arousing mass support and of pressuring hesitant governments. In the coming years, the local conflicts most likely to escalate into major wars will be those, as in Bosnia and the Caucasus, along the fault lines between civilizations. The next world war, if there is one, will be a war between civilizations.

THE WEST VERSUS THE REST

THE WEST IS NOW at an extraordinary peak of power in relation to other civilizations. Its superpower opponent has disappeared from the map. Military conflict among Western states is unthinkable, and Western military power is unrivaled. Apart from Japan, the West faces no economic challenge. It dominates international political and security institutions and with Japan international economic institutions. Global political and security issues are effectively settled by a directorate of the United States, Britain and France, world economic issues by a directorate of the United States, Germany and Japan, all of which maintain extraordinarily close relations with each other to the exclusion of lesser and largely non-Western countries. Decisions made at the U.N. Security Council or in the International Monetary Fund that reflect the interests of the West are presented to the world as reflecting the desires of the world community. The very phrase "the world community" has become the euphemistic collective noun (replacing "the Free World") to give global legitimacy to actions reflecting the interests of the United States and other Western powers.[4] Through the IMF and other international economic institutions, the West promotes its economic interests and imposes on other nations the economic policies it thinks appropriate. In any poll of non-Western peoples, the IMF undoubtedly would win the support of finance ministers and a few others, but get an overwhelmingly unfavorable rating from just about everyone else, who would agree with Georgy Arbatov's characterization of IMF officials as "neo-Bolsheviks who love expropriating other people's money, imposing undemocratic and alien rules of economic and political conduct and stifling economic freedom."

---

[4] Almost invariably Western leaders claim they are acting on behalf of "the world community." One minor lapse occurred during the run-up to the Gulf War. In an interview on "Good Morning America," Dec. 21, 1990, British Prime Minister John Major referred to the actions "the West" was taking against Saddam Hussein. He quickly corrected himself and subsequently referred to "the world community." He was, however, right when he erred.

Western domination of the U.N. Security Council and its decisions, tempered only by occasional abstention by China, produced U.N. legitimation of the West's use of force to drive Iraq out of Kuwait and its elimination of Iraq's sophisticated weapons and capacity to produce such weapons. It also produced the quite unprecedented action by the United States, Britain and France in getting the Security Council to demand that Libya hand over the Pan Am 103 bombing suspects and then to impose sanctions when Libya refused. After defeating the largest Arab army, the West did not hesitate to throw its weight around in the Arab world. The West in effect is using international institutions, military power and economic resources to run the world in ways that will maintain Western predominance, protect Western interests and promote Western political and economic values.

The very phrase "world community" has become a euphemism to give legitimacy to the actions of the West.

That at least is the way in which non-Westerners see the new world, and there is a significant element of truth in their view. Differences in power and struggles for military, economic and institutional power are thus one source of conflict between the West and other civilizations. Differences in culture, that is basic values and beliefs, are a second source of conflict. V. S. Naipaul has argued that Western civilization is the "universal civilization" that "fits all men." At a superficial level much of Western culture has indeed permeated the rest of the world. At a more basic level, however, Western concepts differ fundamentally from those prevalent in other civilizations. Western ideas of individualism, liberalism, constitutionalism, human rights, equality, liberty, the rule of law, democracy, free markets, the separation of church and state, often have little resonance in Islamic, Confucian, Japanese, Hindu, Buddhist or Orthodox cultures. Western efforts to propagate such ideas produce instead a reaction against "human rights imperialism" and a reaffirmation of indigenous values, as can be seen in the support for religious fundamentalism by the younger generation in non-Western cultures. The very notion that there could be a "universal civilization" is a Western idea, directly at odds with the particularism of most Asian societies and their emphasis on what distinguishes one people from another. Indeed, the author of a review of 100 comparative studies of values in different societies concluded that "the values that are most important in the West are least important worldwide."[5] In the political realm, of course, these differences are most manifest in the efforts of the United States and other Western powers to induce other peoples to adopt Western ideas concerning democracy and human rights. Modern democratic government originated in the West. When it has developed in non-Western societies it has usually been the product of Western colonialism or imposition.

The central axis of world politics in the future is likely to be, in Kishore Mahbubani's phrase, the conflict between "the West and the Rest" and the responses of non-Western civilizations to Western power and values.[6] Those responses generally take one or a combination of three forms. At one extreme, non-Western states can, like Burma and North Korea, attempt to pursue a course of isolation, to insulate their societies from penetration or "corruption" by the West, and, in effect, to opt out of participation in the Western-dominated global community. The costs of this course, however, are high, and few states have pursued it exclusively. A second alternative, the equivalent of "band-wagoning" in international relations theory, is to attempt to join the West and accept its values and institutions. The third alternative is to attempt to "balance" the West by developing economic and military power and cooperating with other non-Western societies against the West, while preserving indigenous values and institutions; in short, to modernize but not to Westernize.

---

[5] Harry C. Triandis, *The New York Times*, Dec. 25, 1990, p. 41, and "Cross-Cultural Studies of Individualism and Collectivism," Nebraska Symposium on Motivation, vol. 37, 1989, pp. 41-133.

[6] Kishore Mahbubani, "The West and the Rest," *The National Interest*, Summer 1992, pp. 3-13.

THE TORN COUNTRIES

IN THE FUTURE, as people differentiate themselves by civilization, countries with large numbers of peoples of different civilizations, such as the Soviet Union and Yugoslavia, are candidates for dismemberment. Some other countries have a fair degree of cultural homogeneity but are divided over whether their society belongs to one civilization or another. These are torn countries. Their leaders typically wish to pursue a bandwagoning strategy and to make their countries members of the West, but the history, culture and traditions of their countries are non-Western. The most obvious and prototypical torn country is Turkey. The late twentieth-century leaders of Turkey have followed in the Attatürk tradition and defined Turkey as a modern, secular, Western nation state. They allied Turkey with the West in NATO and in the Gulf War; they applied for membership in the European Community. At the same time, however, elements in Turkish society have supported an Islamic revival and have argued that Turkey is basically a Middle Eastern Muslim society. In addition, while the elite of Turkey has defined Turkey as a Western society, the elite of the West refuses to accept Turkey as such. Turkey will not become a member of the European Community, and the real reason, as President Özal said, "is that we are Muslim and they are Christian and they don't say that." Having rejected Mecca, and then being rejected by Brussels, where does Turkey look? Tashkent may be the answer. The end of the Soviet Union gives Turkey the opportunity to become the leader of a revived Turkic civilization involving seven countries from the borders of Greece to those of China. Encouraged by the West, Turkey is making strenuous efforts to carve out this new identity for itself.

During the past decade Mexico has assumed a position somewhat similar to that of Turkey. Just as Turkey abandoned its historic opposition to Europe and attempted to join Europe, Mexico has stopped defining itself by its opposition to the United States and is instead attempting to imitate the United States and to join it in the North American Free Trade Area. Mexican leaders are engaged in the great task of redefining Mexican identity and have introduced fundamental economic reforms that eventually will lead to fundamental political change. In 1991 a top adviser to President Carlos Salinas de Gortari described at length to me all the changes the Salinas government was making. When he finished, I remarked: "That's most impressive. It seems to me that basically you want to change Mexico from a Latin American country into a North American country." He looked at me with surprise and exclaimed: "Exactly! That's precisely what we are trying to do, but of course we could never say so publicly." As his remark indicates, in Mexico as in Turkey, significant elements in society resist the redefinition of their country's identity. In Turkey, European-oriented leaders have to make gestures to Islam (Özal's pilgrimage to Mecca); so also Mexico's North American-oriented leaders have to make gestures to those who hold Mexico to be a Latin American country (Salinas' Ibero-American Guadalajara summit).

Historically Turkey has been the most profoundly torn country. For the United States, Mexico is the most immediate torn country. Globally the most important torn country is Russia. The question of whether Russia is part of the West or the leader of a distinct Slavic-Orthodox civilization has been a recurring one in Russian history. That issue was obscured by the communist victory in Russia, which imported a Western ideology, adapted it to Russian conditions and then challenged the West in the name of that ideology. The dominance of communism shut off the historic debate over Westernization versus Russification. With communism discredited Russians once again face that question.

President Yeltsin is adopting Western principles and goals and seeking to make Russia a "normal" country and a part of the West. Yet both the Russian elite and the Russian public are divided on this issue. Among the more moderate dissenters, Sergei Stankevich argues that Russia should reject the "Atlanticist" course, which would lead it "to become European, to become a part of the world economy in rapid and organized fashion, to become the eighth member of the Seven, and to put particular emphasis on Germany and the United States as the two dominant members of the Atlantic alliance." While also rejecting an exclusively Eurasian policy, Stankevich nonetheless argues that Russia should give priority to the protection of Russians in other countries, emphasize its Turkic and Muslim connections, and promote "an appreciable redistribution of our resources, our options, our ties, and our interests in favor of Asia, of the eastern direction." People of this persuasion criticize Yeltsin for subordinating Russia's interests to those of the West, for reducing Russian military strength, for failing to support traditional friends such as Serbia, and for pushing economic and political reform in ways injurious to the Russian people. Indicative of this trend is the new popularity of the ideas of Petr Savitsky, who in the 1920s argued that Russia was a unique Eurasian civilization.[7] More extreme dissidents voice much more blatantly nationalist, anti-Western and anti-Semitic views, and urge Russia to redevelop its military strength and to establish closer ties with China and Muslim countries. The people of Russia are as divided as the elite. An opinion survey in European Russia in the spring of 1992 revealed that 40 percent of the public had positive attitudes toward the West and 36 percent had negative attitudes. As it has been for much of its history, Russia in the early 1990s is truly a torn country.

To redefine its civilization identity, a torn country must meet three requirements. First, its political and economic elite has to be generally supportive of and enthusiastic about this move. Second, its public has to be willing to acquiesce in the redefinition. Third, the dominant groups in the recipient civilization have to be willing to embrace the convert. All three requirements in large part exist with respect to Mexico. The first two in large part exist with respect to Turkey. It is not clear that any of them exist with respect to Russia's joining the West. The conflict between liberal democracy and Marxism-Leninism was between ideologies which, despite their major differences, ostensibly shared ultimate goals of freedom, equality and prosperity. A traditional, authoritarian, nationalist Russia could have quite different goals. A Western democrat could carry on an intellectual debate with a Soviet Marxist. It would be virtually impossible for him to do that with a Russian traditionalist. If, as the Russians stop behaving like Marxists, they reject liberal democracy and begin behaving like Russians but not like Westerners, the relations between Russia and the West could again become distant and conflictual.[8]

---

[7] Sergei Stankevich, "Russia in Search of Itself," *The National Interest*, Summer 1992, pp. 47-51; Daniel Schneider, "A Russian Movement Rejects Western Tilt," *Christian Science Monitor*, Feb. 5, 1993, pp. 5-7.

THE CONFUCIAN-ISLAMIC CONNECTION

THE OBSTACLES TO non-Western countries joining the West vary considerably. They are least for Latin American and East European countries. They are greater for the Orthodox countries of the former Soviet Union. They are still greater for Muslim, Confucian, Hindu and Buddhist societies. Japan has established a unique position for itself as an associate member of the West: it is in the West in some respects but clearly not of the West in important dimensions. Those countries that for reason of culture and power do not wish to, or cannot, join the West compete with the West by developing their own economic, military and political power. They do this by promoting their internal development and by cooperating with other non-Western countries. The most prominent form of this cooperation is the Confucian-Islamic connection that has emerged to challenge Western interests, values and power.

Almost without exception, Western countries are reducing their military power; under Yeltsin's leadership so also is Russia. China, North Korea and several Middle Eastern states, however, are significantly expanding their military capabilities. They are doing this by the import of arms from Western and non-Western sources and by the development of indigenous arms industries. One result is the emergence of what Charles Krauthammer has called "Weapon

---

[8] Owen Harries has pointed out that Australia is trying (unwisely in his view) to become a torn country in reverse. Although it has been a full member not only of the West but also of the ABCA military and intelligence core of the West, its current leaders are in effect proposing that it defect from the West, redefine itself as an Asian country and cultivate close ties with its neighbors. Australia's future, they argue, is with the dynamic economies of East Asia. But, as I have suggested, close economic cooperation normally requires a common cultural base. In addition, none of the three conditions necessary for a torn country to join another civilization is likely to exist in Australia's case.

States," and the Weapon States are not Western states. Another result is the redefinition of arms control, which is a Western concept and a Western goal. During the Cold War the primary purpose of arms control was to establish a stable military balance between the United States and its allies and the Soviet Union and its allies. In the post-Cold War world the primary objective of arms control is to prevent the development by non-Western societies of military capabilities that could threaten Western interests. The West attempts to do this through international agreements, economic pressure and controls on the transfer of arms and weapons technologies.

The conflict between the West and the Confucian-Islamic states focuses largely, although not exclusively, on nuclear, chemical and biological weapons, ballistic missiles and other sophisticated means for delivering them, and the guidance, intelligence and other electronic capabilities for achieving that goal. The West promotes nonproliferation as a universal norm and nonproliferation treaties and inspections as means of realizing that norm. It also threatens a variety of sanctions against those who promote the spread of sophisticated weapons and proposes some benefits for those who do not. The attention of the West focuses, naturally, on nations that are actually or potentially hostile to the West.

The non-Western nations, on the other hand, assert their right to acquire and to deploy whatever weapons they think necessary for their security. They also have absorbed, to the full, the truth of the response of the Indian defense minister when asked what lesson he learned from the Gulf War: "Don't fight the United States unless you have nuclear weapons." Nuclear weapons, chemical weapons and missiles are viewed, probably erroneously, as the potential equalizer of superior Western conventional power. China, of course, already has nuclear weapons; Pakistan and India have the capability to deploy them. North Korea, Iran, Iraq, Libya and Algeria appear to be attempting to acquire them. A top Iranian official has declared that all Muslim states should acquire nuclear weapons, and in 1988 the president of

A Confucian-Islamic connection has emerged to challenge Western interests, values and power.

Iran reportedly issued a directive calling for development of "offensive and defensive chemical, biological and radiological weapons."

Centrally important to the development of counter-West military capabilities is the sustained expansion of China's military power and its means to create military power. Buoyed by spectacular economic development, China is rapidly increasing its military spending and vigorously moving forward with the modernization of its armed forces. It is purchasing weapons from the former Soviet states; it is developing long-range missiles; in 1992 it tested a one-megaton nuclear device. It is developing power-projection capabilities, acquiring aerial refueling technology, and trying to purchase an aircraft carrier. Its military buildup and assertion of sovereignty over the South China Sea are provoking a multilateral regional arms race in East Asia. China is also a major exporter of arms and weapons technology. It has exported materials to Libya and Iraq that could be used to manufacture nuclear weapons and nerve gas. It has helped Algeria build a reactor suitable for nuclear weapons research and production. China has sold to Iran nuclear technology that American officials believe could only be used to create weapons and apparently has shipped components of 300-mile-range missiles to Pakistan. North Korea has had a nuclear weapons program under way for some while and has sold advanced missiles and missile technology to Syria and Iran. The flow of weapons and weapons technology is generally from East Asia to the Middle East. There is, however, some movement in the reverse direction; China has received Stinger missiles from Pakistan.

A Confucian-Islamic military connection has thus come into being, designed to promote acquisition by its members of the weapons and weapons technologies needed to counter the military power of the West. It may or may not last. At present, however, it is, as Dave McCurdy has said, "a renegades' mutual support pact, run by the proliferators and their backers." A new form of arms competition is thus occurring between Islamic-Confucian states and the West. In an old-fashioned arms race, each side developed its own arms to balance or to achieve superiority against the other side. In this new form of arms competition, one side is developing its arms and the other side is attempting not to balance but to limit and prevent that arms build-up while at the same time reducing its own military capabilities.

IMPLICATIONS FOR THE WEST

THIS ARTICLE DOES not argue that civilization identities will replace all other identities, that nation states will disappear, that each civilization will become a single coherent political entity, that groups within a civilization will not conflict with and even fight each other. This paper does set forth the hypotheses that differences between civilizations are real and important; civilization-consciousness is increasing; conflict between civilizations will supplant ideological and other forms of conflict as the dominant global form of conflict; international relations, historically a game played out within Western civilization, will increasingly be de-Westernized and become a game in which non-Western civilizations are actors and not simply objects; successful political, security and economic international institutions are more likely to develop within civilizations than across civilizations; conflicts between groups in different civilizations will be more frequent, more sustained and more violent than conflicts between groups in the same civilization; violent conflicts between groups in different civilizations are the most likely and most dangerous source of escalation that could lead to global wars; the paramount axis of world politics will be the relations between "the West and the Rest"; the elites in some torn non-Western countries will try to make their countries part of the West, but in most cases face major obstacles to accomplishing this; a central focus of conflict for the immediate future will be between the West and several Islamic-Confucian states.

This is not to advocate the desirability of conflicts between civilizations. It is to set forth descriptive hypotheses as to what the future may be like. If these are plausible hypotheses, however, it is necessary to consider their implications for Western policy. These implications should be divided between short-term advantage and long-term accommodation. In the short term it is clearly in the interest of the West to promote greater cooperation and unity within its own civilization, particularly between its European and North American components; to incorporate into the West societies in Eastern Europe and Latin America whose cultures are close to those of the West; to promote and maintain cooperative relations with Russia and Japan; to prevent escalation of local inter-civilization conflicts into major inter-civilization wars; to limit the expansion of the military strength of Confucian and Islamic states; to moderate the reduction of Western military capabilities and maintain military superiority in East and Southwest Asia; to exploit differences and conflicts among Confucian and Islamic states; to support in other civilizations groups sympathetic to Western values and interests; to strengthen international institutions that reflect and legitimate Western interests and values and to promote the involvement of non-Western states in those institutions.

In the longer term other measures would be called for. Western civilization is both Western and modern. Non-Western civilizations have attempted to become modern without becoming Western. To date only Japan has fully succeeded in this quest. Non-Western civilizations will continue to attempt to acquire the wealth, technology, skills, machines and weapons that are part of being modern. They will also attempt to reconcile this modernity with their traditional culture and values. Their economic and military strength relative to the West will increase. Hence the West will increasingly have to accommodate these non-Western modern civilizations whose power approaches that of the West but whose values and interests differ significantly from those of the West. This will require the West to maintain the economic and military power necessary to protect its interests in relation to these civilizations. It will also, however, require the West to develop a more profound understanding of the basic religious and philosophical assumptions underlying other civilizations and the ways in which people in those civilizations see their interests. It will require an effort to identify elements of commonality between Western and other civilizations. For the relevant future, there will be no universal civilization, but instead a world of different civilizations, each of which will have to learn to coexist with the others. ◆

The Importance of Space in Violent Ethno-religious Strife

DAVID C. RAPOPORT

The political and moral dimensions of space are rarely understood to be crucial political ingredients. This essay challenges that assumption by examining ethno-religious conflicts since the 1880s. It argues that the massive spatial configurations occasioned by the collapse of great empires created a powerful impuetus for a steady expansion of the number of ethno-religious conflicts. Space decisively shaped features of the struggles, namely length, intensity, and purpose. Finally, space influenced international responses to internal ethno-religious struggles making proximity more important than power in decisions to intervene.

> We have continued in possession of the very land which gave us birth. Sprung from its very soil we are able to address our city by the very names which we apply to our nearest kin; for we alone of all the Hellenes have the right to call our city at once nurse and fatherland and mother.
>
> Isocrates, 'Oration on Autochthony'

> Should the people of America divide themselves into three or four nations...like most other *bordering* nations they would be either involved in disputes and war or live in constant apprehension of them.
>
> John Jay, *Federalist Paper*, No.V (original emphasis)

The first notable discussion[1] of space and conflict appears in *The Federalist*, a good starting point for the subject still.[2] Peoples in neighbouring sovereign states, the authors argue, are natural enemies, but their animosities diminish when they possess territorial autonomies within a single state. The concept of federalism eventually emerges from these premises, and federal institutions are the basis for the most successful solutions of ethno-religious' conflict devised for democratic states.[4]

The Federalist emphasizes physical geography, that is the nature of the terrain, the distances within and between states, issues especially important to the new American successor state. But the volume does not offer a general account of how new state boundaries alter existing social relations. And it ignores the potentiality of space to generate moral feeling, as in cases where sites are considered sacred, and/or when a community appeals to the principle of national self-determination, a *right* to a space it can control.[5]

These dimensions of space (that is, physical, political and moral geography) influence the course, conduct, and consequences of ethno-religious violence. This article will discuss patterns since the 1880s, highlighting the period since 1945 to focus on several questions: what is the relationship between the massive redrawing of political boundaries occasioned by the fall of great empires and the steady expansion of the number of these conflicts? How do claims for space influence special features of ethno-religious strife within states? What are the international responses to domestic struggles for space? This third question is the most significant one.

Redrawing Political Boundaries

The conventional wisdom that the world-wide profusion of ethno-religious conflicts is a peculiar post-Cold-War phenomenon is false, for the number grew steadily throughout the past century. Ethno-religious concerns inspired over half of the violent struggles within states from 1945 to 1960. The proportion increased to three-quarters from 1960 to 1990; and while the collapse of the Soviet world after 1990 produced new, vivid instances, it did not greatly alter the gross statistics.[6]

No reliable count of such conflicts in the world before 1945 exists; my own estimate is that between 1875 and 1914 15 per cent were ethno-religious, a percentage increasing to around 25 between 1918 and 1945. The figures for the two periods before and after 1945 demonstrate constant, albeit uneven, growth.[7]

Simultaneously, a similar expansion in the number of states occurred. Nearly five times as many states exist today as in the 1880s, representing a dramatic restructuring of political space.[8] The overwhelming majority of new states are successor states deriving either from great land empires whose territories were all contiguous (Austro-Hungarian, Ottoman and Soviet) or from sea empires with scattered overseas territories (British, French, Belgian, Dutch and Portuguese).

Obviously, violent ethno-religious struggles are not restricted to successor states. Since the 1960s, Spain (Basques and Catalans), France (Corsicans), Thailand (Pattenis), Ethiopia (Eritreans) and Afghanistan (fundamentalists, traditionalists, and tribal elements) have experienced the phenomena; and none is a successor state. Imperial disintegrations, moreover, do not always leave legacies of ethno-religious strife; the collapse of the Spanish American empire in the early nineteenth century

---

David C. Rapoport, Center for the Study of Religion, UCLA

Nationalism & Ethnic Politics, Vol.2, No.2, Summer 1996, pp.258–85
PUBLISHED BY FRANK CASS, LONDON

NOTICE: THIS MATERIAL [illegible]
BY COPYRIGHT LAW (TITLE 17 U.S CODE)

WHARTON REPROGRAPHICS

3

produced many unstable successor states, but few have endured ethno-religious violence, and all those experiences have been short-lived.[9]

But successor states are more vulnerable than others because spatial reconfigurations exacerbate existing ethno-religious tensions.[10] Since the boundaries of earlier states and their successors rarely coincide, new possibilities for enormous and traumatic consequences may develop. Minorities may become majorities, and majorities minorities. The birth of the Irish Free State produced a Catholic minority in Northern Ireland, and, should the island host one state, Protestants will become the minority.

Mediating elements may disappear. Without the British, for example, relationships between Cypriot Greeks and Cypriot Turks changed dramatically. Internal balances and alliances can be destroyed. In the former Yugoslavia, for example, Croatians combined with Slovenians to check Serbian influence; the Slovenian secession, therefore, badly strained relationships between all remaining ethno-religious elements.

A people once united in a single state may suddenly find themselves scattered. The dissolution of the former Yugoslavia fragmented the Serbian community, putting fractions in different states where all but one became a minority. The disappearance of the Soviet state left 25 million Russians (17 per cent of the Russian population) living in successor states outside Russia, potentially beleaguered minorities in territories where their language was once dominant. The Soviet collapse also left one-third of the Armenian population outside Armenia. The most vexing issue was the fate of those in Nagorno-Karabakh in Azerbaijan, troublesome enough when all were part of a larger state, but a disaster in the new circumstance.[11] When the Ottoman empire collapsed its Kurdish population was divided between the successor states of Turkey, Syria and Iraq.

As the Kurds everywhere then lived near international borders, their loyalties became suspect. A similar situation confronts a fragmented Serbian population. The special vulnerability of border peoples is a perennial problem perhaps as old the state itself, as the Bible's description of Pharaoh's decision to enslave Israel, a border people living in Goshen, suggests: 'These people...have become so numerous and strong, that they are a threat to us...if war should break out they might take arms against us.'[12]

New state boundaries create enormous uncertainties about the future of groups, uncertainties which may be cruelly aggravated if linked to memories of injustices and atrocities and to fears of how groups will exploit their new strength or protect new vulnerabilities. These anxieties will be especially intense at moments when a new state is created or anticipated, for that is when violence and fresh succession efforts occur most often.[13] Those who witnessed the births of India and Pakistan should not have been surprised by the violence and new secessionist claims which occurred after the initially peaceful dissolutions of the Soviet Union and Yugoslavia. Ireland, Sudan, Angola, Mozambique, Rwanda, Burundi, Zaire, Cyprus, Iraq, Myanmar (Burma), Cambodia and the Philippines, to name a few, saw similar struggles, which began during the battle for independence or shortly afterwards, and have persisted intermittently ever since.[14]

Space and the Dynamics of Internal Conflict

Space shapes distinguishing features of wars: their length, intensity and purpose. Most ethno-religious struggles are a species of internal wars, wars *within* rather than *between* states; and, characteristically, internal wars are longer than international ones, six times longer in the twentieth century. Today internal wars persist six times longer than they did a century ago. One reason is that since 1945 most internal wars have been ethno-religious, and as the proportion of such struggles keeps growing so does the length of internal wars.[15] Normally, length and intensity (casualties and numbers participating) are inversely related. But internal wars with ethno-religious issues witness more attempts at genocide and more frequent abuses of human rights.[16] Wars end either by capitulation or negotiation, but neither method works as well in wars with ethno-religious dimensions.[17] In the unusual event of a decisive victory, attempts at genocide occur in its wake.[18]

Why are ethno-religious struggles so savage and intractable? 'Identity' questions, conventional wisdom suggests, command our deepest emotions, and significantly space is so closely connected with identity that whole lands and particular sites often appear sacred. Yet there is more to the problem.[19] International wars involving such questions are briefer and more restrained than internal wars based on them. Also negotiated settlements of international conflicts, whatever the issues, are more durable than those of internal wars.

Space has other pertinent features too. International wars normally leave each side with its own state intact; and because secession movements rarely succeed, the protagonists must continue sharing the same political space. To protect the separate character of the parties to an internal settlement, institutional checks and immunities, 'time-buying' arrangements, can be created; but time does not always heal and such 'solutions' could become recipes for indecision and, therefore, renewed tensions.

Historically, contending parties in an internal war usually aim to capture or hold the government in order to dominate all the state's available space. Before the French and the American Revolution, ethno-religious struggles also had this feature, and as Hutu and Tutsi violence in Rwanda and Burundi illustrate, they sometimes still do. But in the past two centuries it is more common to believe that a people needs a land of its own to be complete, a doctrine which encourages rebels to try to redraw state boundaries. The doctrine can lead to territorial autonomy claims within an existing state, to self-determination claims for a sovereign state or to *irredentist* (*enosis*) aspirations for union with related group(s) in other existing sovereign state(s).[20] Compared with the last two alternatives, territorial autonomy is least threatening to the internal order and no problem for the international one.

Obviously, claims for space are influenced by the fact that a community already is concentrated in space.[21] But the doctrine that every people has a right to 'national self-determination' seems more significant; indeed, today no principle of legitimacy is more compelling or more explosive, because its existence makes every state vulnerable to being challenged by those who feel themselves a people apart from the one existing states represent. And there is no way to realize this principle fully without dissolving most existing states, perhaps 90 per cent of them.[22]

The very possibility that this claim can be made means that quarrels originating in other, less far-reaching issues may be transformed. A quarrel over the status of English as a national language in Sri Lanka, for example, moved inevitably towards secession after blood was spilled. Violence, as Clausewitz knew, has its own logic, making the stakes of a struggle higher and higher in order to justify the ever-increasing flow of sacrifices made, one reason why wars are more difficult to stop than to start. Note how Ferrero describes the dynamics of the French Reign of Terror:

> The Jacobins did not spill all that blood because they believed in popular sovereignty as a religious truth; they tried to believe in popular sovereignty as a religious truth because their fear made them spill all that blood.[23]

The doctrine that every people has a right to a space adequate to govern itself, as its genesis in the American and the French Revolution suggests, is intimately connected with the idea of democracy. But democracy is also associated with formal procedures which sometimes clash with the idea of self-determination. Democracies are uniquely required to determine majorities and minorities, and that necessity often exacerbates ethno-religious tensions. When majorities and minorities are identified in ethno-religious terms they tend to become permanent entities; and permanent majorities and minorities, *The Federalist* observed, are extremely prone to violence.[24]

A century later, John Stuart Mill wrote that, as the great empires were most often multi-ethnic realms, the spread of democratic principles to their successor states would exacerbate existing tensions.[25] Canada provides the most recent example for his case. When it finally broke all legal ties with Britain in 1982 to adopt a new constitution, the issue of Quebec, the only province with a French-speaking majority, became even more troublesome than it had been before. English speaking provinces refused to support Quebec's status as an entity with powers different from their own.

International Ramifications of Internal Struggles for Space

> The chief, the overwhelming problem of diplomacy seems to be weak states, ...'weak' because they are industrially backward and at present politically incompetent...to prevent outbreaks of internal violence.
>
> Walter Lippmann, *The Stakes of Diplomacy*

Violence within states is an important, perhaps the most important, precipitant of wars between them. Matters provocative enough to generate war will certainly stimulate less costly acts, interventions to stop, moderate and encourage internal violence. But which states intervene and why? What is the relationship between strife within a state and the general structure of international tensions? The physical properties of space, namely proximity to neighbours is an important – perhaps decisive – concern, one ignored too often. In this section I shall also pay special attention to neighbouring states possessing elements who share an identity with the rebels. I conclude by noting some moral features of space and with observations on the related new doctrine for intervention, *moralpolitik*.

Physical Space: the Importance of Proximity

According to conventional wisdom, the most powerful states intervene most often. No one can deny that they are the most active international parties or that their interventions have had the most dramatic, bloody and far-reaching consequences; witness the genesis of World War I, World War II, the Korean and the Vietnamese War. Yet in each of these cases the importance of neighbourhood may be shown too, and one can make a good *a priori* case for proximity as being the decisive circumstance. Neighbours feel the most immediate effects in the form of refugees and potential changes in regional power balances; neighbours, of course, are the best placed to act. Finally, rebels will want to involve neighbours because porous borders are critical to success.

Ethno-religious violence creates an additional impetus for intervention by neighbours. The boundaries of ethno-religious communities generally overlap those of neighbouring states, creating persistent waves of sympathies and antagonism transcending borders. The efforts of Armenians in Nagorno-Karabakh to break loose from Azerbaijan brought the neighbouring state of Armenia into the fray. Nearly a million Azeri (a

Turkic people) became refugees, and after much hesitation, Turkey blockaded Armenia. Many Kurds, and all their enemies, believe that the creation of Kurdistan requires Iraq, Iran, Turkey and Syria to give up territories. Turkish Cypriots believe that *enosis* (the aspiration of Greek Cypriots to join Greece) threatens them, and Turkey agrees. The aim of Islamic fundamentalists to unify the Islamic world provokes anxiety among states with significant Muslim populations.

Related activities provide another reason for seeing neighbourhood as most important. Richardson and Wilkinson's study of wars between 1820 and 1945 concluded that 94 per cent occurred between neighbours, and that by far the most important element associated with all wars studied was contiguity. Even the number of frontiers a state possessed related directly to the number of wars it fought.[26] While contiguous sovereign states may share many important political and cultural traditions, their interests keep conflicting, which is why *The Federalist* argued that it is much more difficult to explain neighbours who never go to war with each other than those who frequently do.

But where does the weight of historical evidence lie concerning neighbourhood versus power? Evidence from the first half of the century is inconclusive. From 1870 to 1914 the global powers intervened again and again, but their territories were so extended that they were neighbours in places where one or the other intervened. Proximity and power then combined to make regional conflicts world concerns, a pattern so striking that in 1914 Lippmann predicted that internal violence would be the century's most persistent *international* issue.[27] Most examples of internal violence in this period did not involve ethno-religious strife, but when that was the issue it proved to be the most dangerous and intractable, especially in the Balkans, an Ottoman legacy on the frontiers of several major powers. Imperial absorption, usually a neighbour's last solution for persistent eruptions which might spread, failed here. Thus in 1908 a search for peace in the Balkans led European powers to place Bosnia under a reluctant Austro-Hungarian jurisdiction, despite the fact that most Bosnians preferred union with Serbia. Bosnians were vulnerable to Serbian intrigues, precipitating a sixth Balkan war and thereby World War I.[28]

Ethno-religious violence became less significant for global tensions after World War I, partly because Western empires had absorbed so much of the world and therefore the opportunities for neighbours to intervene in troubled territories declined. Important interventions occurred, and Nazi *irredentist* demands upon three neighbouring Austro-Hungarian and German successor states (Austria, Czechoslovakia and Poland) were important ingredients in the genesis of World War II.

After 1945, when so many empires dissolved, the earlier link between power and neighbourhood was ruptured, a fact which makes it easier to assess the comparative importance of power versus neighbourhood. When one compares neighbouring states with Cold War protagonists and then Western with Soviet metropolitan cores *after* their respective empires dissolved, the greater significance of neighbourhood becomes apparent.

Cold War Protagonists and Neighbouring States. One Cold War 'truism' is that participants were deeply involved in shaping power struggles everywhere. Yet the dangers of over-extension were understood also, a phenomenon commentators were less likely to recognize. Both sides avoided unnecessary, marginal, dangerous conflicts, marginal because they had only local implications and dangerous because they were savage and intractable. The fact is that Cold War adversaries were much more successful than global powers had been earlier in avoiding involvements in ethno-religious conflicts, which is one reason why the World War which twice before developed out of such tensions did not occur again.[29]

The restraint of Cold War adversaries was especially striking when the issue was secession. (They were less successful regarding *irredentist* claims, a matter I return to later.) In some 40 secessionist struggles during the Cold War, neighbouring states opposed each other in 'most instances' but Cold War protagonists favoured different sides only four times.[30]

In those four cases support was always contingent on the larger Cold War objective. The Soviet treaty with India encouraged the latter, for example, to aid Bangladesh to secede from Pakistan; but the Soviet purpose was to gain India as an ally. In the strategic Horn of Africa, the Soviets indirectly gave military aid to Eritrean rebels; but when Ethiopia joined the Marxist world the Soviets deserted the Eritreans to give Ethiopia arms and training. Iraqi Kurds in the 1970s received valuable Soviet diplomatic support; but when the possibility of a Soviet–Iraqi alliance occurred, the Kurds paid the price and 'settled' for autonomy. After World War II the Soviets supported Azeri rebels and a Kurdish republic (*Mahabad*) in Iran, the latter a Soviet objective since the 1920s, but under American pressure, the Soviets abandoned both groups. To counter Soviet influence in Nicaragua, the US encouraged the Miskito Indians to revolt, a rare example of ethnic insurrection in Latin America.[31] But when the Soviet-supported Sandinista government was forced out, the US abandoned the Miskitos. Similarly, American help for the Tibetan rebels (the Khampas) vanished when Cold War tensions with China dissipated in the 1970s.[32]

In other instances where Cold War protagonists were on different sides, important global interests were at stake but secession was not. Angola, thus, was the gateway to a Cold War prize, the rich neighbouring state of South Africa, and the Soviets and Americans supported different ethnic factions wanting to capture or hold the state. Several Lebanese conflicts involved competing Cold War adversaries, but no Cold War party supported secession.[33] Afghan rebels received enormous international aid. But secession was not an issue; the Soviet invasion to support an illegal government gave the US and several Muslim states (Pakistan, Iran, Saudi Arabia and Egypt), who agree on very little, a common cause.[34]

Although aid to rebels was almost invariably covert, the statistics demonstrate that neighbouring states more frequently and more persistently intervened in secessionist struggles. In one important respect, Cold War adversaries and neighbouring states have similar policies: aid levels were usually subject to more important concerns. Spanish Basque terrorists found sanctuary on French soil where Basques lived until Franco died. The Irish Republic initially gave the IRA valuable political and logistic support in 1969, when the latest 'troubles' began. As the Republic became confident that the Catholics in Northern Ireland would survive, her policies became less provocative to London.[35]

The Basque and the Irish examples are not unique.[36] Uganda's strong ethno-religious ties to rebels in the Sudan were a reason for logistic support, but Uganda itself needed money too, and that was more important. Thus when the Ugandan dictator Idi Amin, himself a Kakwa (a Sudanese rebel tribe), decided that Libya's Colonel Gadaffi was more generous than Israel, Gadaffi's enemy and supporter of the Sudanese rebels, Amin turned his back on his kinsmen, even though rebel troops from the Sudan and the Sudanese issue helped him rise to power in 1971.[37] Bonds between neighbouring states and rebels are stronger than those between rebels and Cold War protagonists because neighbours are often kinfolk, but one would be foolish to reckon that kin ties are always decisive. Ethiopian logistic support for Sudanese rebels reflected Christian solidarity and an anti-Islamic sentiment. But the Sudan also had been sustaining Ethiopian rebels (Eritrean), and that was more important to Ethiopia whose help waxed and waned whenever the Sudan's support for the Eritreans did.

Ethno-religious conflict rarely begins by demanding secession; that goal becomes more likely after blood has been shed. As sucession becomes more explicit, neighbours become more uneasy for two reasons: many have their own secession potentialities, and the integrity of existing states is a cardinal principle of the international order. India's policy towards Sri Lanka's troubles provides a most dramatic example. India initially supported the largely Hindu Tamils against the Buddhist Sinhalese majority in Sri Lanka, a policy which also reflected anger at the latter's decision to allow Pakistan to use its airfields during the 1971 Indo-Pakistan War. When Tamils opted for secession, New Delhi reconsidered, fearing that Indian Tamils might try to secede too. Indian troops, therefore, helped to enforce the July 1987 agreement Sri Lanka had made for greater Tamil autonomy. (After a thousand Indian lives were lost troops were withdrawn in 1990 with little to show for their efforts. Tamils also assassinated Rajiv Ghandi, India's premier, for 'treason'.)

Many other and more obscure examples exist. To get a border readjustment, Malaysia gave diplomatic support to the grievances of fellow Muslims (Pattenis) rebelling in Thailand. When the Patteni goal became secession, that support vanished. The Philippines began training troops to take territory disputed with Malaysia; Malaysia responded by providing funds and military aid to the Moros, fellow Muslims in the Philippines. After the Moros became secessionists they were abandoned, but not before the Filipinos understood the point Malaysia was making.

Israel aided rebels to weaken states, mostly bordering ones, still in a technical state of war with Israel. She encouraged ethnic rebellions in Iraq and Lebanon, and went further afield to supply military training and equipment to Sudanese rebels. The same policy applied to governments too. Israel gave Ethiopia, a Christian finger in the dike against a Muslim sea, help against Eritrean rebels supported by several Muslim states. When the Soviets embraced Ethiopia Israeli aid ceased, but after the Soviets reversed their policy, Israel returned.

A government will normally get more generous support than rebels will, especially from neighbours with territories adjacent to those controlled by that government. Rebels in control of those territories will receive the same benefit, except that aid to rebels is normally covert. It is embarrassing to defy the norms of sovereignty openly, demonstrating again (though does one need another demonstration?) that 'hypocrisy is the tribute vice always pays to virtue'.[38]

But the secrecy patterns display significant variations. Pakistan conceals support for Sikhs and Nagas in India; when fellow Muslims are involved as in the Kashmir, Pakistani presence is more visible, more extensive, and less a matter of *realpolitik*. But governments rarely recognize a secessionist group's claim to be a state.[39]

Rebels bent on secession are strategically weak; they need their neighbours more than neighbours need them. Over and over again the same pattern recurs. States will encourage a neighbour's secessionists when that neighbour seeks to stimulate rebellion in those states. Some rebels are so weak strategically that they become playthings of the neighbourhood. 'Orphans of the universe', the Kurds dispersed in four states always suffer from foreign intrigues. A most gruesome experience occurred in the 1970s when Iran armed Iraqi Kurds to push Iraq to an agreement about the disputed Shatt-al Arab territories. When Iran got what it wanted (1975) the Kurds were left dangling between the ferocious revenge of Saddam Hussein and a Turkey unwilling to accept Kurdish refugees for fear that they might stir up Turkish Kurds.[40] Foreign involvements constantly exacerbated divisions among Kurds themselves, a matter Palestinians understand well.

Powerful proof of the significance of neighbouring states emerges when one compares them with more powerful, but more distant states. Take the case of Cyprus. When Greek troops helped Cypriots seeking union with Greece to scrap a constitution guaranteeing the Turkish minority veto power, Turkey invaded the island in 1974 to establish a second state composed of Cypriot Turks, one which only Turkey recognizes. Britain, though a guarrantor of the country's original constitution, was no longer a neighbour and could be brushed aside.

The Cyprus problem has produced discussions, resolutions and actions by several international bodies (the UN, NATO and the Conference of Islamic States) and by a number of individual states. But Greece and Turkey remain the key actors, and fear of war with Turkey led Greece to dampen Greek Cypriot enthusiasm for *enosis*, which Greece initially fanned when Britain ruled the island. Intermittent American efforts to create a single state again have always been frustrated, leaving the US to settle for any situation which would keep both Greece and Turkey in NATO. But acceptable situations are not necessarily satisfactory ones, necessitating the presence of a UN peacekeeping force ever since.

The tangled strife of Lebanon during the 1970s and the 1980s, a strife precipitated by the influx of Palestinian refugees from neighbouring states, illustrates similar patterns. Both Cold War powers had interests in the outcome and were sometimes directly on the scene. Still, the Soviets had minimal influence; they were allied with both the Syrians and the Palestine Liberation Organization (PLO), but despite Soviet objections, the Syrians continually hampered and weakened the PLO. The US put troops in Lebanon, but made a humiliating withdrawal after the destruction of the Marine barracks in 1983. The importance of neighbours was paramount. Israel and Syria were consistently involved, having the greatest interests and the most direct physical access. Both sent troops and also armed sectarian militias, playing them off one against the other. Other regional powers (Saudi Arabia, Iran and Iraq) intervened using distinctive methods, yielding different results.

Secession did not gain much support in Lebanon, a result which would have upset most neighbouring states. One neighbour, Israel, did favour partition but it could live with a competent, moderate government. The Lebanon experience finally demonstrates that when a region is already torn by international conflict, ethnic discord in one state could drag in most of the neighbours, complicating and intensifying issues.

Irredentist claims have even greater potential for upsetting international arrangements. Once violence occurs, pressures on foreign states to become involved seem irresistible, which may explain why Cold War adversaries found it more difficult to avoid these conflicts than those involving secession. In Korea and Vietnam two states claimed to represent one nation and the conflicts became hot wars in which major Cold War adversaries participated.[41] Withdrawal was particularly difficult here; all four local parties were recognized as states and each was deeply committed to the Cold War, ideological struggle, circumstances which solidified alliances and made major Cold War adversaries neighbours. Korea and Vietnam did not produce the war everyone feared most, an outcome which probably would have happened in a war between the two German states, because they were much more important and at the centre of the two alliances.

To be sustained, *irredentist* claims require foreign support from states with the ethnic elements necessary to consummate the union. Greece's active encouragement induced the Cypriot revolt. Would there have been trouble in the 1930s in Czechoslovakia if Germany had not agitated? In this respect, the case of the rebels in the Italian Tyrol after World War II is revealing. When Austria did not (could not?) support them, the rebellion collapsed. An equally revealing situation is present in Ireland. As long as the Irish constitution claims jurisdiction over Ulster there may be no possibility that Protestants and Catholics will find a way to end the violence.

Triangular or quadrangular relationships inherent in *irredentist* claims complicate and intensify international tension.[42] Several times Greece and Turkey almost went to war over Cyprus. The decision at Munich on the Sudetenland was a crucial step towards World War II. This special tension may explain why *irredentist* claims are made less often than those for self-determination.[43]

Former Imperial States; Neighbours and Non-neighbours. The contention that space is more important than power becomes quite compelling when states shorn of their empires are compared. Their distinctive geographies help to explain why the post-imperial intervention policies of Western and Soviet metropolitan cores were so different.

Western states remained powerful, retaining large investments in former territories where many of their nationals remained. Before 1945 those circumstances in the wake of internal violence would have produced extensive military interventions and often imperial annexations. After 1945 direct military interventions in the successor states were limited to rescue any nationals threatened, and if there was another purpose it was to bolster existing governments with diplomatic aid, limited military help and other forms of assistance. We may note British support for Nigeria against Biafra, US assistance for the Philippine government against the Moros, and French help for governments in Chad, Gabon, Togo and Cameroon. The only exception might be Belgian military assistance to Katangan secessionists in the Congo (Zaire) in 1960. But the Belgians were ambivalent, and later (1978) Belgian and French troops helped to put down a similar rebellion in the same province.[44]

Post-imperial policies of Western states did vary sometimes. France was by far the most active former imperial power, believing herself to have special interests in all French-speaking African states (including those once under Belgian administration), a view those states generally accepted.[45] French troops intervened in the former Belgian territories in 1990 and 1993 for the Rwanda government and against Tutsi rebellions. In 1994 France intervened again, ostensibly for humanitarian purposes, so gruesome had Hutu government atrocities against the Tutsi become. The French created 'safe havens', limiting for a time the victory of the Tutsi-dominated Patriotic Front.

The consequences of supporting the integrity of successor states were forseeable, for unstable fearful governments often violate human rights and abuse particular ethnic elements. A French-supported government in Rwanda initiated attempts at genocide in 1994, whereupon the French public seriously questioned the wisdom of a policy always aimed at strengthening governments. More profound consequences occurred when France stood behind an Algerian government which denied Islamic parties opportunities to gain electoral success. A serious terrorist campaign began in July 1995, presumably by Islamic elements determined to break French resolution. Algeria and France are neighbouring states, and the significance of this geography becomes clearer when we describe the disintegration of the Soviet empire.

The post-imperial record of the Soviet metropolitian core provides a striking contrast with those of the Western empires, particularly with respect to secession. In retrospect, it seems paradoxical that, although Russia was the only imperial power to give up her hold voluntarily, it now is, and will probably remain, an active intervening power. A recent estimate from the Russian General Staff claims that there may be as many as 30 conflicts in former Soviet territories which are, or about to be, violent, a number which could grow to 70 soon.[46] Even before the Chechneya rebellion began (1994), some 40,000 Russian troops were engaged in various ways, usually as peacekeepers.

There are two significant spatial differences between the Western empires and the Soviet one. The former were sea empires, while the Soviets had a land empire, one with wholly contiguous territories resembling its Ottoman and Austro-Hungarian predecessors. Russia's size is unprecedented too: no previous metropolitan core remained a neighbour of virtually all its successor states.

Space helps to explain the issues contended too. Boundary disputes are inevitable legacies of imperial disintegrations. Western successor states disputed boundaries with each other but not with the metropolitan core. Although most boundary disputes occur in the outlying states, Russia also has many boundary problems and some are serious, namely, the dispute with Ukraine over the Crimea. The seas provide indisputable boundaries between metropolitan cores and colonial territories. Land empires are not so fortunate.[47]

Land boundaries are generally more permeable than maritime ones. In the Soviet case boundaries of successor states were often created initially by Communist fiat to sever ties with historic or natural boundaries, a fact of great potential significance.[48] The Soviet dissolution left 25 million Russians in surrounding successor states. To force them out some successor states have adopted discrimination policies which, in effect, mean that no Russian government can survive if it abandons those in the 'near abroad'.[49]

The size and location of populations which had migrated from metropolitan cores may be a troubling problem. In the sea empires the most bitter and prolonged violence occurred in lands such as Algeria and Kenya where sizeable European populations resided. Once independence was achieved, the hostility to Europeans was so great that most immediately left. In the case of Algeria a considerable number of Algerians on the 'wrong side' in the war for independence fled to France too, making it even more difficult to separate the politics of the two states. The 'Irish problem' has been so persistent because the island contained a high concentration of Scottish (Protestant) settlers. When the British bought Cyprus, many Turks remained, and after it achieved independence, Turkish Cypriots looked to Turkey for help against Greek Cypriots and received it.

The Russian term for the world occupied by other Soviet successor states – the 'near abroad' – indicates an intention to make the Commonwealth of Independent States (CIS) more cohesive than any political association left by Western empires (such as the British Commonwealth of Nations or the Annual Franco–African Summit). To induce the reluctant to join and to secure mutual defence agreements that would keep her troops on the scene, Russia first backed rebels and then governments in Moldava, Georgia and Azerbaijan.[50] Russian populations in successor states are major elements in promoting a larger union, activities which increase Russia's interest in their security.

Russia' manipulates forces in unstable neighbouring successor states, partly to check other neighbouring states' efforts. Violent secession movements make *irredentist* claims to join other successor states, namely, Ossetians in Georgia wish to reunite with Russia and elements in Moldava want to rejoin Romania.[51] Inevitably, apprehensions of Russian intentions emerge in other states of the former Soviet world which seek third-party security guarantees. Thus the efforts of former Warsaw Pact states to join NATO, which inevitably provoke Russian suspicions.

Changes in spatial relations in Central Asia and the Caucasus are creating new circumstances, unknown either to the Soviets or the Tsars. Since 1991 five new Turkic republics (Azerbaijan, Kazakhstan, Kyrgyzstan, Turkmenistan and Uzbekistan) have been looking to Turkey for inspiration and support. This pressure could reorient Turkish policy from Europe towards Central Asia for the first time since the early Ottoman Empire. Even though Turkic refugees from the Caucasus and Central Asia traditionally found asylum among the Ottomans, and the early republic had a vision of uniting all Turkic peoples in one national state, preoccupation with Europe continued. Recent manifestations of this potential reorientation are Turkey's blockade of Armenia to help Azerbaijan and the presence of Turkish volunteers in Chechneya.[52] If Turkey becomes active in Central Asia, can Russia be passive?[53]

The establishment of unstable successor states on the former Soviet Union's southern flank made Iran a power in the region for the first time in a century. But the dynamics of the two periods are quite different. Earlier, Iran defended itself against Russian imperial expansion; now the political appeal of Islamic fundamentalism puts Russia, with a significant Muslim population, on the defensive.

Beyond the problem of unstable neighbouring successor states, another serious possibility exists. The extraordinary difficulties in coping with the uprising in Chechnya makes one wonder whether Russia can keep its own secessionists elements at bay;[54] disintegration threats in the metropolitan cores of imperial states are more common than most realize.[55]

Will either the left or the right revive Russian imperial ambitions? A resolution in March 1996 from a Communist-dominated legislature declared that the dissolution of the Soviet Union was null and void and the response of non-Communist executive was to create union between Belarus and Russia. One should not forget that there is an interesting precedent in Russian history. Lenin armed the Soviet Union with a new doctrine for a multi-ethnic state to regain seceding Tsarist territories.[56] And there are non-Russian parallels: the German imperial revival, for example, after the Versailles humiliation. Against this, one could cite Russia's own precarious internal unity, her voluntary decision to sever links with most of the other republics,[57] the facts that empires rarely revive, and anti-imperial sentiment everywhere else is so strong. But Russia's intentions may be irrelevant, because a power with so many unstable neighbours must worry immensely about its own security and thus become active in their affairs.

Another imponderable element is that Russia, unlike its Western counterparts, is experiencing a radical new understanding of its identity. If the Ottoman experience is a useful analogue, where the Armenians, Kurds and others provide grim testimony, the process could have serious consequences for minorities.[58] Permeable land boundaries and numerous refugees from ethnic turmoil elsewhere gave Russia large non-Russian populations. Difficult Russian economic circumstances compound the problem, as enterprising refugee elements excite considerable envy.[59]

The catastrophic speed of this redefining process has no historic parallel. An empire which took four centuries to build collapsed virtually overnight without a major war and without time to come to terms with the likelihood of its dissolution. Significantly, during the Soviet period Russians, unlike their Western counterparts, did not see themselves as an imperial people so much as one engaged with other nationalities in common efforts to spread world revolution; and the unexpected hostilities and humiliations have been traumatic.

Moral Space: the Development of Moralpolitik

When after World War II, the principle of national self-determination became universally accepted, a crucial dividing line between the two halves of the century materialized. In declaring that self-determination was a prerequisite for all other rights, a UN Resolution created the basis for *moralpolitik*.[60] *Realpolitik* is a politics based on state interests and *moralpolitik* is divorced substantially from them.

Moralpolitik had striking implications for three important matters: conquest, imperialism and human rights. The most remarkable consequence, certainly the one least noticed, was that annexation via conquest virtually disappeared from the international scene. Four conquests did occur in 50 years, but each was associated with special circumstances. South Vietnam was the only state extinguished by conquest; the struggle began against French imperial rule and North Vietnam made *irredentist* claims. India conquered a Portuguese territory (Goa, 1961) and international recognition of the act suggested that Western imperial domains were fair game. However, when Indonesia annexed another Portuguese possession (East Timor, 1975), the act remained unrecognized. Although the occupation of East Timor has not yet been reversed, indigenous resistance provoked savage responses, discrediting Indonesia's reputation with regard to human rights. Israel did not attempt to annex the West Bank, and her efforts to annex the Golan Heights and Jerusalem were not internationally recognized, which means that the return of the Golan is a condition of peace with Syria and a final Palestinian settlement may require a retreat on Jerusalem.

More conspicuous than the barrier to new conquests, and closely related to it, was the implication of national self-determination and *moralpolitik* for existing imperial states. Empires were seen as products of historical conquests and secessionist movements within them attracted several kinds of sympathetic support from abroad, always from the global South and from the Soviets, and sometimes even from Western elements too.[61]

On the other hand, the metropolitan cores of Western empires and all imperial successor states were recognized as national states. Secessionists in those entities received little international credibility; in Europe the Basques and the Corsicans were largely ignored.[62] Nearly 50 secessionist movements occurred in non-European successor states (1960–90), but only one succeeded. Bangladesh broke away from Pakistan (1971) after a newly-signed defence treaty with the Soviets enabled India to invade Pakistani territory.[63]

Violent dismemberment of successor states may be impossible without significant outside help and that happens so rarely that states with the slimmest prospect for survival do endure. The new African states maintained themselves,[64] just as did the most problematic new states of the nineteenth century, those in Latin America.[65]

I have already noted that *irredentist* claims are made less often than those for self-determination, partly because they require the active support of kindred foreign states and are more likely to produce international war. But it seems also relevant that states aspiring to absorb groups in other states evoke the image of conquerors to the international world. Finally, states with *irredentist* claims are usually multi-ethnic themselves and expanding one ethnic element may seriously transform domestic politics. The most successful and least bloody example is the takeover of Goa, an absorption which could not alter domestic Indian balances and, of course, one which 'reversed' an imperial condition.

When states act collectively in international organizations, opposition to secessionist insurgency seems strongest, because in such circumstances states must make their reasons public and the integrity of existing states is a principle all accept. The UN even used troops to prevent Katanga (Shaba) from seceding from the Congo (Zaire) in 1960–64. The record of collective bodies with strong Islamic influence is less clear. The Islamic Conference (43 states) endorsed 'the right of the fraternal people of the Turkish Muslim Community of Cyprus', to be heard at all international forums(1976). But only Turkey recognizes the Turkish Cypriot state.[66] The League of Arab States (torn by three contradictory concerns: Arab, Islamic and national influences) experiences the same conflicts that divide member successor states. Moral support to Muslim rebels in the Philippines and Ethiopia makes it the only regional organization to encourage secessionists in non-imperial contexts. But the League never endorsed the secessionist aim as such, urged rebels to seek a peaceful solution within the framework of the existing state.

and invariably condemned secessionist rebellions in Arab states.

Moralpolitik, most often connected with international organizations, is conspicuous in commitments to human rights, particularly in violent situations where peacekeeping, medical and famine relief efforts occur. The antecedents were the post-World War I treaties, which established Austro-Hungarian successor states and simultaneously recognized that the self-determination principle often created serious problems for minorities. The League of Nations pledge to protect minorities proved ineffective, but the UN which retained the principle hoped to enforce it, though little was done during the Cold War.[67]

After the Cold War, beliefs in the value of collective action for such activity intensified. Nearly two-thirds of UN peacekeeping operations occurred in this brief period, and for the first time major powers (except China) sent troops. When the belligerents themselves desire assistance and intervention is limited to narrow tasks, intervention may be effective as UN involvements in Cyprus (1964–?) and Somalia (1992–95) indicate. But it is difficult to expand the consensus to deal with peacemaking; note the experiences in Somalia and the former Yugoslavia.[68]

Shortly after the former Yugoslavia dissolved, the UN *Declaration of the Rights of Persons Belonging to National or Ethnic, Religious, or Linguistic Minorities (1992)* gave *moralpolitik* its most explicit formal ground. The *Declaration* is an enormous breach in the principle of non-interference, providing a right, quite possibly a duty, to intervene collectively. It is too early to assess its import,[69] but the momentum for action, let alone successful intervention, requires both a greater consensus on defining minority rights and their pre-eminent importance than is presently available. We need, too, a better understanding of the ultimate effect that well-intentioned, poorly-supported interventions have on the directing organization's credibility.

Conclusion

Violence within states always disturbs relations between them, and ethno-religious violence, perhaps the most troubling form, has spread steadily since the late nineteenth century. After World War II half of the internal struggles were ethno-religious; by the 1960s ethno-religious violence outstripped all others put together, a fact which Cold War preoccupations obscured.

As the great multi-ethnic empires collapsed, each successor state sought to link itself with a nation, a course which usually produced both violence and failure. The spread of democratic principles may have contributed to these disasters too, and no legitimate alternatives to the democratic form exist now.

New boundaries seriously disturb pre-existing ethno-religious relations, reversing majority–minority relationships, eliminating traditional mediating elements, fragmenting peoples, and suddenly casting some in vulnerable border areas. Uncertainties about the future are maximized at the moment of creation, making violence most likely then.

The resulting conflicts are long, intense, and intermittent. Space, the object of the struggles, helps to explain these features. To have a space of one's own is a crucial ingredient in identity; but participants in internal struggles have to cohabit a territory afterwards.

Distances between states help to explain who intervenes, why, how and the probability of success. When great powers had extensive territories, they were more likely to be each other's neighbours, and internal violence, especially that animated by ethno-religious issues, could spark global conflict. The creation of new states created new neighbours, changed the identity of the interveners, and *reduced* the likelihood that a conflict would drag others outside the region in. Because its spatial dimensions are so unusual, Russia is the most active intervening power today. But as long as its territories do not impinge on those of other major powers, it is unlikely that global conflicts will be precipitated.

Each new space created made the world *larger*. A counter-intuitive argument to be sure, but only because one's political imagination is imprisoned by the metaphor of a global village which technology keeps shrinking daily. But in a world which grows larger, most individual struggles become more local. We may be more aware that an ethno-religious conflict exists, but that does not confer a greater will to act *decisively*. Indeed, the relationship may be an inverse one. Even in the Cold War the major protagonists and the international community as a whole were reluctant to involve themselves in ethno-religious strife, except when that strife was directly associated with major concerns in the Cold War or with Western imperialism, and especially when both elements were understood as combined, as in the Palestinian and the South African case.

Proximity makes some states much more crucial than conventional estimates of power would have us believe. Neighbours have substantial interests in the outcome of violence next door, interests which may lead them to stimulate and aggravate tensions, or alternatively they may try to put the fires out, fearing dismemberment precedents and wars unrelated to their interests.

Violent ethno-religious movements have less manoeuvring room than states do; time and again rebels are abandoned or given aid to continue but never enough to win. This game can be played badly, and neighbouring states may be dragged into war especially in areas where pre-existing hostility exists, such as the Middle East or the Balkans. Still the continuing

Balkan problem is quite instructive. Disagreement among UN participants probably prolonged the violence and strained venerable alliances. But the Balkan predicament historically always threatened to and occasionally did, precipitate war between the major outside intervening powers, only a remote possibility now. Tensions may still escalate,[20] but since 1945 ethnic bond claims have rarely been strong enough to provoke war between neighbouring states. Greece made a humiliating retreat when Turkey showed that it would risk war rather than abandon Turkish Cypriots. (Would the outcome have been different if Greece and Turkey had been on opposing sides in the Cold War?)

The importance given to the 'right to space' obviously shapes intervention patterns in decisive ways too. Conquest, hitherto a common solution for internal violence, became intolerable, and as empires became illegitimate, secessionist hopes grew. But because the international world did not describe successor states as imperial regimes secessionists later had less success. Violence did not cease, and a new framework for intervention developed: 'human rights', 'humanitarian assistance' and 'prevention of genocide', interventions which are usually collective.

Moralpolitik developed, a politics divorced substantially from state interests as conventionally defined, a politics created by a world which had grown larger and attractive most to those who need to salve their conscience and are geographically remote from the scene of the conflict. But the imperatives of *realpolitik* have not disappeared and *moralpolitik* creates issues it may not resolve; trying war criminals in the Balkan conflict could, for example, be an insuperable obstacle to a durable peace.

Moralpolitik was given its most comprehensive basis in the 1992 UN *Declaration of the Rights of Persons*, which created a legal but not a political capacity to act. Earlier experiences in Lebanon, Somalia and Yugoslavia suggest that by the time a problem is defined in humanitarian terms it may have already become virtually insoluble by limited interventions. The UN failed its first major test after the 1992 *Declaration* in the 1994 Rwanda crisis, and the first state to act (France) had substantial interests invested. A related test was the UN (the USA in this instance) intervention to restore democracy and a respect for human rights in Haiti (October 1994). The extraordinary indecision and last-minute compromises associated with the action may signify the absence of substantial questions of interest.[21] Simultaneously, a significantly different response to Iraqi mobilization on Kuwaiti borders in summer 1990 reflected the presence of substantial interests and a different principle.[22]

In 1991 two events which seemed to signify changing attitudes towards violent secession opened a Pandora's box.[23] Eritrea was recognized. Then Germany, Austria and Hungary recognized Croatia's right to secede from the former Yugoslavia, a precedent followed by some other NATO powers, who ignored the European Community's own demands that its standards for minority rights be met. These were decisive steps both in the dissolution of Ethiopia and the former Yugoslavia.[74]

But that is not the whole story. Ethiopia unlawfully incorporated Eritrea, and the case, therefore, may never be a precedent. Yugoslavia's tragedy may teach us that premature recognitions are agonies to avoid, not precedents to follow, and the more recent disappointments of the Kurds suggest that this second reading is more probable. Kurds were exhorted to rise during the Gulf War; and to protect them from Saddam Hussein's revenge, UN Resolution 688 (April 1992) created a safe zone protected by the victorious allies. The Kurds seized the opportunity to establish a quasi-state to put secession on the table again. But every relevant power seems determined to maintain Iraq's integrity; dismemberment anxieties are very much alive still.[75]

My concern has been with intervention in specific states; and in approaching the issue that way, one might reasonably conclude that the problem is serious for the international world as a whole, but 'manageable'. But what about the cumulative, world-wide effect of strife in so many countries? This is a subject for another study and a less optimistic one. Perhaps the most important such results are the appearance of refugees or the mass movement of people. In November 1993 (*before* the most recent floods from Rwanda and the former Yugoslavia) one out of every 130 persons in the world was a refugee. The long-run consequences of this ever-increasing flow affects us all, albeit unequally. Weaker states often carry the heaviest burden in Asia and Africa. In Europe sudden migrations from the Balkans exacerbate internal problems occasioned by earlier population movements, providing fuel for growing right-wing movements and violence. Traditional rights of political asylum are being severely restricted, and does anyone think the matter will end there?

NOTES

1. Early versions of this article were prepared for the Institute of Global Conflict and Co-operation, University of California (15 Jan. 1995), UCLA Religion Colloquium (11 May 1995) and the APSA Annual Confermce (30 Aug. 1995). I am grateful for the comments of Donald Rothchild, David Lake, Fred Wehling and Sandra Halperin there. Michael Barkun, Marita Kaw, Ibrahim Karawan, Jennifer Pournelle, Avivah Rapoport, Barbara Rapoport, Khachig Tololyan and Joan Witte all made useful suggestions.
2. Space now often refers to astronomical systems. Here I return to a more classical notion; space refers to geographic expanses on earth where social systems are located. Political scientists rarely pay attention to the importance of space. Nicholas J. Spykman is perhaps the most distinguished American exception. His 'Geographic Objectives in Foreign Policy I' (with Abbie Rollins), *American Political Science Review*, Vol.XXXIII, No.3 (June 1939), pp.391–410 and 'Geographic Objectives in Foreign Policy II', Ibid., No.4 (Aug.1939), pp.591–604 develop the theoretical bases of geopolitics for his later books, *America's Strategy in World Politics* (New York: Harcourt Brace, 1942) and *The Geography of the Peace* (New York: Harcourt Brace, 1944). For an excellent discussion of Spykman see David Wilkinson, 'Spykman and Geopolitics', in Ciro E. Zoppo and Charles Zorgbibe (eds.), *On Geopolitics: Classical and Nuclear* (Dordrecht: Nijhoff, 1985) pp.77–129.
3. The statistical data used here were collected by others as 'ethno-religious' violence, meaning violence by groups which are ethnic or religious or have features of both. The term nationalist is often used to describe the same groups and Walker Connor has persuasively argued that there is no sense in distinguishing ethnic and national groupings and, hence, the title of his perceptive study, *EthnoNationalism: The Quest for Understanding* (Princeton: Princeton University Press, 1994)

   Elsewhere I argue that religious violence has special properties; see my 'Sacred Terror: a Contemporary Case from Islam',in Walter Reich, *Origins of Terrorism: Psychologies, Theologies, States of Mind* (New York: Cambridge University Press) pp.103–31, and 'Some General Observations on Religion and Violence', *Terrorism and Political Violence*, Vol.3, No.3 (Autumn 1991), pp.118–40, reprinted in Mark Juergensmeyer, *Violence and the Sacred in the Modern World* (London: Cass, 1991), pp.118–30, and 'Comparing Militant Fundamentalist Movements and Groups', in Martin Marty and Scott Appleby (eds.), *Fundamentalisms and the State* (Chicago: University of Chicago Press, 1993).
4. *The Federalist* suggests that cultural differences between neighbouring states are a consequence not the original cause of their antagonisms. Although federalism is a good mechanism to restrain ethno-religious conflict, it is not clear that *The Federalist* had ethno-religious conflict in mind, and there are special circumstances in such cases. Quebec is the only French-speaking Canadian province, and therefore has to be treated differently from all other provinces.
5. The neglect of the moral dimension of space appears ironic because the American *Declaration of Independence* introduced the principle of national self-determination into the political world. But the authors of *The Federalist* restricted themselves to the immediate question of why the Constitution should be ratified.
6. Roy Licklider, 'Negotiated Settlements in Civil Wars Since 1945', *APSR*, Vol.89, No.3 (Sept. 1995), p.685. I have no numbers for subsequent events; the number probably has grown but the percentage increase cannot be great.
7. Significantly, Walter Lippmann's important discussion of relationships between domestic violence and international interventions before 1914 offers only one example of ethno-religious conflict: the struggle in the Balkans; see *Stakes of Diplomacy* (New York: Macmillan, 1915).
8. Complicated imperial jurisdictions make it difficult to assess the number of states in the 1880s but the estimate of 42 for 1883 by *The World Cyclopedia of Universal Knowledge* (New York: World Manufacturing Co., 1884), p.619 seems reasonable.
9. In Latin America native peoples normally define their struggles in revolutionary not ethnic terms. Thus, for example, the *Sendero Luminoso* in Peru and the struggle largely by Mayans which began in the state of Chiapas in Mexico, Jan. 1994 are claims to remake the whole system not claims for autonomy or secession, typical objectives of an ethnic community and the concern of this essay. There were of course exceptions; see later in the text. The Latin American experience indicates that Spain virtually eliminated ethnic differences. Ironically, at home Spain was less successful as Basque and Catalonian revolts show.
10. The distinction between successor states and others is not always made. See, e.g., David Lake, 'Ethnic Conflict and International Intervention', *IGCC Policy Brief*, No.3, Mar. 1995
11. The Armenian campaign to transfer Nagorno-Karabakh during Khrushchev's leadership was the only case before *perestroika* where an ethno-religious group mobilized to seek a border change. See Julian Birch, 'Border Disputes and Disputed Borders in the Soviet Federal

employment of all fundamental human rights'. General Assembly Resolution 637A/7, cited by A. Michalska, 'Self-Determination in International Law', in W. Twinning (ed.), *Issues of Self-Determination* (Aberdeen: Aberdeen University Press, 1991), p.79.

61. The two different attitudes towards secession are expressed in UN documents and Resolutions. The 'Self Determination of Peoples' is mentioned in the Charter, and its use to justify secessionist movements in empires was most firmly expressed in the *Declaration of the Granting of Independence to Colonial Countries and Peoples* of 1960. The 'Friendly Relations' *Declaration* of 1970 rejects secession from 'independent states'. The Geneva Protocols (1977), additions to the 1949 Geneva Conventions, make a similar distinction; Protocol I (art. 1, para. 4) treats 'wars of national liberation' as though they were international conflicts while Protocol II reserves to existing states full rights of self-defence against other insurgencies. See John Dugard, 'International Terrorism and the Just War' in D. C.Rapoport and Yonah Alexander (eds.), *The Morality of Terrorism* New York: Columbia University Press, 2nd edn, 1989), pp.77–98.

    Ironically, this 'double standard', often called anti-Western, was first applied by European states aiding rebels in the Ottoman empire.

62. Rebels in the Soviet sphere also gathered little international support, though in eastern Europe they were crushed so quickly the issue never fully materialized.

63. Richard Sisson and Leo Rose, *War and Secession: Pakistan, India and the Creation of Bangladesh* (Berkeley: University of California Press, 1990), pp.198–201. The treaty was necessary to reassure Indians that China would not intervene.

64. Robert H. Jackson and Carl G. Rosberg, 'Why Africa's Weak States Persist: the Empirical and the Juridical in Statehood', *World Politics*, Vol.35 (Oct. 1982), p.20.

65. Colombia was the only Latin American state dismembered, when it could not prevent Panama from seceding (1903). After the Colombian Senate refused to ratify the canal treaty, the US encouraged a revolt and took military and political (i.e., a 'premature' recognition) steps to protect the insurrection.

66. Naomi Black, 'The Cyprus Conflict', in Astri Suhrke and Lela Noble (eds.), *Ethnic Conflict in International Relations* (New York: Praeger, 1977), p.60.

67. L.P. Mair, *The Protection of Minorities: the Working and Scope of the Minorities Treaties under the League of Nations* (London: Christophers, 1928) and John Coakley, 'Approaches to the Resolution of Ethnic Conflict', *International Political Science Review*, Vol.15, No.3 (July 1994), pp.297–314. A 1966 UN Declaration revived interest in the general problem, culminating in the 1992 Declaration.

68. Collective efforts at peacekeeping by regional groups have shown similar problems. Note, for example, ill-fated efforts by the Arab League in Lebanon (1976) and the Organization of African States in Chad (1981).

69. For a good discussion of the 1992 Declaration see Koen Koch,'The International Community and Forms of Intervention in the Fields of Minority Rights Protection', in Culbertson and Leibowitz, *Minorities*.

70. A dispute over the status of Macedonia, for example, could precipitate a broader war between the four neighbouring states (Bulgaria, Albania, Greece and Turkey) with vital interests there. Greece and Turkey will be opposed on a second issue, more serious than their Cyprus dispute.

71. On successive days different *New York Times* reporters interpreted the concern of the Clinton administration to emphasize or downplay human rights abuses in Haiti as indications of whether an invasion was imminent. See 18 and 19 July 1994, pp.A-5, A-4.

72. No issue seems to galvanize the UN more than an invasion of a duly-constituted state, and it is significant that of the three UN *military* interventions (Congo, Korea and Kuwait) two were against invasions.

73. 'In the wake of the astounding events of the last three years, one can detect a weakening of the existing taboo against secession, indeed the signs of an emerging paradigm shift whereby secession will no longer be treated as unthinkable by the international system.' Heraclides, 'Secession, Self-Determination and Non-Intervention', p.399.

74. See Misha Glenny, *The Fall of Yugoslavia* (London: Penguin,1992), pp.12–14. John Zametica, *The Yugoslav Conflict*, Adelphi Paper No.270 (London: IISS/Brassey's 1992), and John Newhouse, 'The Diplomatic Round', *New Yorker*, 24 Aug. 1992, p.63. Although the Croatians appeared to count on recognition, they also overestimated both the ability and the will of a sympathetic German government to do more.

75. Note also that no state has recognized the recent secession of Somaliland from Somalia. See Rayika Omaar, 'Somaliland: One Thorn Bush at a Time', *Current History* (May 1994), pp.232–6.

4

WHY AMERICA THINKS IT HAS TO RUN THE WORLD

by BENJAMIN SCHWARZ

The Cold War is over, and America is staggering under a colossal debt and an accumulation of frightening social problems. Yet it continues to spend billions to protect Germany and Japan—two rich nations whose freedom is in no apparent danger. Why? Here is the answer that the foreign-policy elite would give if it dared to speak frankly about the delicate matter of American efforts to assert international economic and political control

THREE years ago, in light of the end of the Cold War, the Clinton Administration undertook a "fundamental reassessment" of America's national-security requirements. But after six months of analysis Administration officials concluded that the defense of U.S. global interests still demanded military spending of more than $1.3 trillion over the following five years and the permanent commitment of more than 200,000 U.S. soldiers in East Asia and Europe—in other words, a strategy remarkably similar to that which America pursued during the Cold War. Moreover, rather than relinquish America's costly and risky responsibilities by dissolving Cold War alliances, defense strategists now plan to expand NATO's responsibilities eastward. Those who call for a more modest plan argue that U.S. strategy seems to be extravagance born of paranoia, or of the defense establishment's anxiety to protect its budget. In fact, given the way the makers of U.S. foreign policy have defined American interests since the late 1940s, these plans are quite prudent. And that is the problem.

If many Americans had been asked ten years ago why U.S. troops were deployed in East Asia and Europe, they would have answered, To keep the Soviets out. They may have wondered, however, why the United States persisted in its strategy long after Japan, South Korea, and Western Europe had become capable of defending themselves. Now that the USSR itself has disappeared, why does Washington continue to insist that U.S. "leadership" in East Asia and Europe is still indispensable?

Ask National Security Council staff members, think-tank analysts, or State Department policy planners about America's globe-girdling security commitments and they will deliver very different answers—ones that have not changed in forty-five years. They will justify the Pax Americana by invoking "the imperative of continued U.S. world leadership," the need to "shape a favorable international environment," "reassurance of allies," and the ongoing need for "stability" and "continuing engagement." Even during the Cold War the "Soviet threat" might not have been mentioned.

The question that all this justification ignores is What, exactly, is "leadership," and why has it been the mantra of foreign-policy cognoscenti for nearly fifty years? What have we been doing around the globe, and why?

Most Americans misunderstand their country's foreign policy. It seems to operate only when "danger" looms—when Iraq invades Kuwait, when Russian "imperialism" threatens to resurface, when China rattles its sabers at Taiwan. Even people who religiously read the newspapers fail to grasp that U.S. foreign policy is far more than simply a series of responses to crises.

For instance, media coverage of recurring tensions on the Korean Peninsula has focused on speculation about North Korea's nuclear program and the prospect of a new Korean war. But when foreign-policy officials and experts discuss the Korean crisis among themselves, they rapidly leave the Koreas behind to focus on the real players in the region: China and Japan. As far as national-security experts are concerned, almost any immediate crisis is subsumed by a larger threat—in this case no less than East Asia's role in the potential collapse of the international economy that U.S. power has sustained since the late 1940s.

It's now an axiom of the U.S. foreign-policy establishment that economic, technological, and demographic changes are making East Asia the world's most dynamic arena, a driving force—increasingly the dominant force—in the international economy. The Pacific Century, we are told ad nauseam, has dawned. This transformation also means a shift in the international distribution of political and military power. In a typical evaluation of East Asia's strategic future the foreign-policy expert Aaron Friedberg states darkly in the journal *International Security*,

> In the long run, it is Asia [rather than Europe] that seems far more likely to be the cockpit of great power conflict. The half millennium during which Europe was the world's primary generator of war (as well as of wealth and knowledge) is coming to a close. But, for better and for worse, Europe's past could be Asia's future.

Friedberg's assertion nicely illustrates the ambivalence with which the U.S. national-security community views East Asia's future. He both prophesies an exhilarating Pacific Century and warns the West that the East may once again be up to no good.

East Asia has never been a terribly successful field for American diplomacy. There are undoubtedly many reasons for this, and surely that shortcoming for which the United States is continually indicted—cultural and historical myopia—has contributed enormously to its failures in the region. Americans have always seen East Asia not for what it is but for what it can do to them or for them: the region is either danger or opportunity—either a new "ground war in Asia" or a new China market. American understanding of Japan, for instance, is, in the words of the historian Bruce Cumings, caught within the conflicting views of Japan as "miracle and menace, docile and aggressive, fragile blossom and Tokyo Rose." As Friedberg's analysis attests, the U.S. foreign-policy community worries that the fragile blossom may again bloom into a Tokyo Rose.

Pacific Century rhetoric usually describes the new era in a "post–Cold War" context. This is misleading, because it starts at the wrong place. First, the shift of economic activity has not been sudden. Even if East Asia rose in the American consciousness just as the Soviet Union receded, to define the economic and geopolitical transformation of East Asia as a post–Cold War phenomenon Americanizes and trivializes a development in international (not American) politics of far greater impact than the Cold War itself. Although Vietnam, China, and North Korea were for forty years able to contain America's Cold War ambition to "roll back" communism, they are proving utterly unable to contain the juggernaut of East Asia's capitalist political economy. Most important, to imply that the end of the Cold War is of primary significance to U.S. policy in East Asia is to wrench that policy out of its most important context and to distort its underlying aims and challenges.

What we think of as the Cold War was merely instrumental in America's larger "Cold War" strategy. In "scaring hell out of the American people," as Senator Arthur Vandenberg said in 1947, the U.S.-Soviet rivalry helped to secure domestic support for Washington's ambition to create a U.S.-dominated world order. That same year one of Vandenberg's colleagues, the fervently anti-Communist Senator Robert Taft, expressed a strong suspicion that the supposed dangers to the nation from the USSR failed to explain America's new foreign policy. He complained that he was "more than a bit tired of having the Russian menace invoked as a reason for doing any—and every—thing that might or might not be desirable or necessary on its own merits." The former Secretary of State Dean Acheson put things in proper perspective: describing how Washington overcame domestic opposition to its internationalist policies in 1950, he recalled in 1954 that at that critical moment the crisis in Korea "came along and saved us."

MAKING THE WORLD SAFE FOR CAPITALISM

A FUNDAMENTAL aim of America's Cold War strategy was to create and maintain what the former Secretary of State James Baker has called "a global liberal economic regime"—a capitalist world order. After the Second World War, American statesmen believed that the United States, standing alone and strong in a world of weary nations, had a remarkable opportunity, as Acheson said, to "grab hold of history and make it conform." American statesmen seized that opportunity by creating a complex strategy to reify Adam Smith's dream. Washington envisioned a world economy in which trade and capital would flow across national boundaries in response to the laws of comparative advantage and supply and demand—an economy in which production and finance would be integrated on a global scale. The constricted national markets that were emerging in the immediate aftermath of the Second World War in Europe and East Asia would be combined, eliminating the inefficiencies of statism and self-sufficiency. Large-scale regional economies would in turn be integrated into an interdependent world economy. U.S. policymakers knew that building this multinational capitalist community required the United States to provide Western Europe and Japan with enormous amounts of economic aid (through such schemes as the Marshall Plan, for Europe, and the Dodge Plan, its equivalent for Japan), so that those areas would not retreat into closed economies. They also knew that an open world economy demanded an even more ambitious American project: transforming international relations.

The greatest danger to U.S. democracy and prosperity came, they believed, not from the Soviet Union but from Germany and Japan, whose potential strength amounted to a sort of Catch-22. Without a flourishing international economy, Under Secretary of State Will Clayton warned in 1947, "our democratic free enterprise system" could not function. As late as 1960 exports accounted for only 3.8 percent and imports for 4.8 percent of GDP. The health of the international economy, in this disputable view, depended on Germany's and Japan's economic revitalization. Germany, if its economy was resurrected, would once again be Europe's most efficient producer and most avid consumer. Its very economic potential, however, made Germany a threat to the other Western European states, which, as the future Secretary of State John Foster Dulles explained to a closed Senate panel in 1949, were "afraid to bring that strong, powerful, highly concentrated group of people into unity with them." Similarly, as Dulles, Acheson, and other policymakers understood, a strong Japan was both necessary for a prosperous international order and intolerable to its neighbors. The problem lay in the inherent contradiction between capitalism and international politics.

Capitalist economies prosper most when labor, technology, and capital are fluid, so that they are driven toward international integration and interdependence. But whereas all states benefit absolutely in an open international economy, some states benefit more than others. In the normal course of world politics, in which states are driven to compete for their security, the relative distribution of power is a country's principal concern, discouraging economic interdependence. Thus 250 years ago the philosopher David Hume bemoaned the lack of economic cooperation among countries, blaming the "narrow malignity and envy of nations, which can never bear to see their neighbors thriving, but continually repine at any new efforts towards industry made by any other nation." In its efforts to ensure the distribution of power in its favor and at the expense of actual or potential rivals, a state will "nationalize"—that is, pursue autarkic policies, practicing capitalism only within its borders or among countries in a trading bloc. This circumscribes both production factors and markets, and thereby fragments an international economy.

A truly global economy is probably impossible to achieve. In fact, as the Princeton political economist Robert Gilpin has said, "what today we call international economic interdependence runs so counter to the great bulk of human experience that only extraordinary changes and novel circumstances could have led to its innovation and triumph over other means of economic exchange." Historically, to secure international capitalism a dominant power must guarantee the security of other states, so that they need not pursue autarkic policies or form trading blocs to improve their relative positions. This suspension of international politics through hegemony has been the fundamental aim of U.S. foreign policy since the 1940s. The real story of that policy is not the thwarting of and triumph over the Soviet threat but the effort to impose an ambitious economic vision on a recalcitrant world.

NURTURING—AND CONSTRAINING—THE FRAGILE BLOSSOM

THE overriding objective of U.S. postwar policy toward East Asia was restoring Japanese economic power. To be sure, this would help to immunize the region against Communist expansion; but Acheson and the other creators of the American Century thought the goal itself vital, regardless of the Soviet threat. As the historians William Borden, Bruce Cumings, Ronald McGlothlen, and Michael Schaller have argued, in attempting to create a global economic system Washington pursued a course designed in essence to restore the Greater East Asian Co-Prosperity Sphere—imperial Japan's regional economy, which the United States had just destroyed in the Second World War. If Japan was to help fuel the world economy, it would have to be, in Acheson's words, "the workshop of Asia." American postwar planners viewed the political economy that had developed in Northeast Asia from about 1900 to 1945 as the region's natural economy—a tripartite system in which Japan, given access to continental markets and raw materials, formed the industrial core and its neighbors formed the economic semi-periphery and periphery. As Japan advanced in the product cycle, climbing up the technological ladder, it spun off its low-technology and low-wage industries to its neighbors.

In the late 1940s and the 1950s the United States essentially restored this system; but the machine could not run by itself. Washington had to ensure that Japan's neighbors would feel secure politically in the regional hierarchy atop which Japan stood. Washington also had to ensure that the hierarchy did not develop—as it had in the past—into a Japanese-led closed economic bloc, which would threaten the world economy. NSC 48, the National Security Council's 1949 blueprint for America's Cold War strategy in East Asia, summed up the chief difficulty facing Washington's economic goals in the region—and around the globe. Starting with the premise that "the economic life of the modern world is geared to expansion," requiring "the establishment of conditions favorable to the export of technology and capital and to a liberal trade policy throughout the world," NSC 48's authors went on to warn that "the complexity of international trade makes it well to bear in mind that such ephemeral matters as national pride and ambition can inhibit or prevent the necessary degree of international cooperation, or the development of a favorable atmosphere and conditions to promote economic expansion."

The distinguished historian and diplomat George Kennan, then the head of the State Department's policy-planning staff, saw only one solution to what he described as the "terrible dilemma" confronting U.S. ambitions in East Asia. Forty-seven years later Washington continues to pursue that course.

WHARTON REPROGRAPHICS

The man in the Baghdad café

Which "civilisation" you belong to matters less than you might think

GOERING, it was said, growled that every time he heard the word culture he reached for his revolver. His hand would ache today. Since the end of the cold war, "culture" has been everywhere—not the opera-house or gallery kind, but the sort that claims to be the basic driving force behind human behaviour. All over the world, scholars and politicians seek to explain economics, politics and diplomacy in terms of "culture-areas" rather than, say, policies or ideas, economic interests, personalities or plain cock-ups.

That's just what they're like

Perhaps the best-known example is the notion that "Asian values" explain the success of the tiger economies of South-East Asia. Other accounts have it that international conflict is—or will be—caused by a clash of civilisations; or that different sorts of business organisation can be explained by how much people in different countries trust one other. These four pages review the varying types of cultural explanation. They conclude that culture is so imprecise and changeable a phenomenon that it explains less than most people realise.

To see how complex the issue is, begin by considering the telling image with which Bernard Lewis opens his history of the Middle East. A man sits at a table in a coffee house in some Middle Eastern city, "drinking a cup of coffee or tea, perhaps smoking a cigarette, reading a newspaper, playing a board game, and listening with half an ear to whatever is coming out of the radio or the television installed in the corner." Undoubtedly Arab, almost certainly Muslim, the man would clearly identify himself as a member of these cultural groups. He would also, if asked, be likely to say that "western culture" was alien, even hostile to them.

Look closer, though, and the cultural contrasts blur. This coffee-house man probably wears western-style clothes—sneakers, jeans, a T-shirt. The chair and table at which he sits, the coffee he drinks, the tobacco he smokes, the newspaper he reads, all are western imports. The radio and television are western inventions. If our relaxing friend is a member of his nation's army, he probably operates western or Soviet weapons and trains according to western standards; if he belongs to the government, both his bureaucratic surroundings and the constitutional trappings of his regime may owe their origins to western influence.

The upshot, for Mr Lewis, is clear enough. "In modern times," he writes, "the dominating factor in the consciousness of most Middle Easterners has been the impact of Europe, later of the West more generally, and the transformation—some would say dislocation—which it has brought." Mr Lewis has put his finger on the most important and least studied aspect of cultural identity: how it changes. It would be wise to keep that in mind during the upsurge of debate about culture that is likely to follow the publication of Samuel Huntington's new book, "The Clash of Civilisations and the Remaking of World Order".

The clash of civilisations

A professor of international politics at Harvard and the chairman of Harvard's Institute for Strategic Planning, Mr Huntington published in 1993, in *Foreign Affairs*, an essay which that quarterly's editors said generated more discussion than any since George Kennan's article (under the by-line "x") which argued in July 1947 for the need to contain the Soviet threat. Henry Kissinger, a former secretary of state, called Mr Huntington's book-length version of the article "one of the most important books... since the end of the cold war."

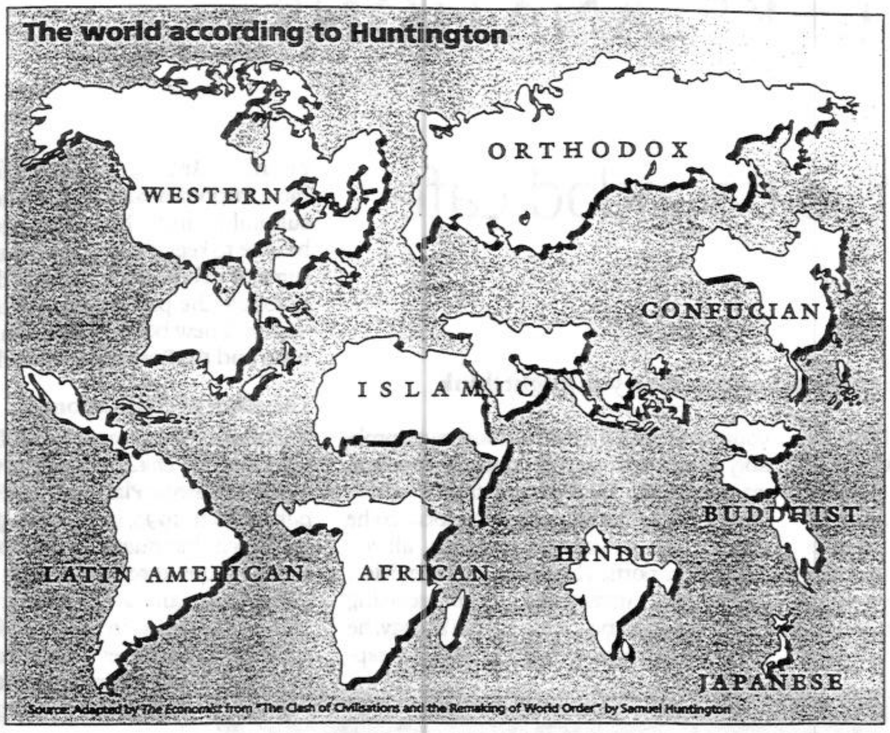
The article, "The Clash of Civilisations?", belied the question-mark in its title by predicting wars of culture. "It is my hypothesis", Mr Huntington wrote, "that the fundamental source of conflict in this new world will not be primarily ideological or primarily economic. The great divisions among humankind and the dominating source of conflict will be cultural."

After the cold war, ideology seemed less important as an organising principle of foreign policy. Culture seemed a plausible candidate to fill the gap. So future wars, Mr Huntington claimed, would occur "between nations and groups of different civilisations"—western, Confucian, Japanese, Islamic, Hindu, Orthodox and Latin American, perhaps African and Buddhist. Their disputes would "dominate global politics" and the battle-lines of the future would followed the fault-lines between these cultures.

No mincing words there, and equally few in his new book:

> Culture and cultural identities... are shaping the patterns of cohesion, disintegration and conflict in the post-cold war world... Global politics is being reconfigured along cultural lines.

Mr Huntington is only one of an increasing



The world according to Huntington

Source: Adapted by *The Economist* from "The Clash of Civilisations and the Remaking of World Order" by Samuel Huntington

number of writers placing stress on the importance of cultural values and institutions in the confusion left in the wake of the cold war. He looked at the influence of culture on international conflict. Three other schools of thought find cultural influences at work in different ways.

- **Culture and the economy.** Perhaps the oldest school holds that cultural values and norms equip people—and, by extension, countries—either poorly or well for economic success. The archetypal modern pronouncement of this view was Max Weber's investigation of the Protestant work ethic. This, he claimed, was the reason why the Protestant parts of Germany and Switzerland were more successful economically than the Catholic areas. In the recent upsurge of interest in issues cultural, a handful of writers have returned to the theme.

It is "values and attitudes—culture", claims Lawrence Harrison, that are "mainly responsible for such phenomena as Latin America's persistent instability and inequity, Taiwan's and Korea's economic 'miracles', and the achievements of the Japanese." Thomas Sowell offers other examples in "Race and Culture: A World View". "A disdain for commerce and industry", he argues, "has ... been common for centuries among the Hispanic elite, both in Spain and in Latin America." Academics, though, have played a relatively small part in this debate: the best-known exponent of the thesis that "Asian values"—a kind of Confucian work ethic—aid economic development has been Singapore's former prime minister, Lee Kuan Yew.

- **Culture as social blueprint.** A second group of analysts has looked at the connections between cultural factors and political systems. Robert Putnam, another Harvard professor, traced Italy's social and political institutions to its "civic culture", or lack thereof. He claimed that, even today, the parts of Italy where democratic institutions are most fully developed are similar to the areas which first began to generate these institutions in the 14th century. His conclusion is that democracy is not something that can be put on like a coat; it is part of a country's social fabric and takes decades, even centuries, to develop.

Francis Fukuyama, of George Mason University, takes a slightly different approach. In a recent book which is not about the end of history, he focuses on one particular social trait, "trust". "A nation's well-being, as well as its ability to compete, is conditioned by a single, pervasive cultural characteristic: the level of trust inherent in the society," he says. Mr Fukuyama argues that "low-trust" societies such as China, France and Italy—where close relations between people do not extend much beyond the family—are poor at generating large, complex social institutions like multinational corporations; so they are at a competitive disadvantage compared with "high-trust" nations such as Germany, Japan and the United States.

- **Culture and decision-making.** The final group of scholars has looked at the way in which cultural assumptions act like blinkers. Politicians from different countries see the same issue in different ways because of their differing cultural backgrounds. Their electorates or nations do, too. As a result, they claim, culture acts as an international barrier. As Ole Elgstrom puts it: "When a Japanese prime minister says that he will 'do his best' to implement a certain policy," Americans applaud a victory but "what the prime minister really mean was 'no'." There are dozens of examples o misperception in international relations, ranging from Japanese-American trade disputes to the misreading of Saddam Hussein's intentions in the weeks before he attacked Kuwait.

What are they talking about?

All of this is intriguing, and much of it is provocative. It has certainly provoked a host of arguments. For example, is Mr Huntington right to lump together all European countries into one culture, though they speak different languages, while separating Spain and Mexico, which speak the same one? Is the Catholic Philippines western or Asian? Or: if it is true (as Mr Fukuyama claims) that the ability to produce multinational firms is vital to economic success, why has "low-trust" China, which has few such companies, grown so fast? And why has yet-more successful "low-trust" South Korea been able to create big firms?

This is nit-picking, of course. But such questions of detail matter because behind them lurks the first of two fundamental doubts that plague all these cultural explanations: how do you define what a culture is?

In their attempts to define what cultures are (and hence what they are talkin about), most "culture" writers rely partly o self definition: cultures are what people think of themselves as part of. In Mr Huntington's words, a civilisation "is the broadest level of identification with which [a person] intensely identifies."

The trouble is that relatively few people identify "intensely" with broad cultural groups. They tend to identify with something narrower: nations or ethnic groups. Europe is a case in point. A poll done last year for the European Commission found that half the people of Britain, Portugal and Greece thought of themselves in purely national terms; so did a third of the Germans, Spaniards and Dutch. And this was in a part of the world where there is an institution—the EU itself—explicitly devoted to the encouragement of "Europeanness".

The same poll found that in every EU country, 70% or more thought of themselves either purely in national terms, or primarily as part of a nation and only secondly as Europeans. Clearly, national loyalty can coexist with wider cultural identification. But, even then, the narrower loyalty can blunt the wider one because national characteristics often are—or at least are often thought to be—peculiar or unique. Seymour Martin Lipset, a sociologist who r cently published a book about nation characteristics in the United States, called i "American Exceptionalism". David

Willetts, a British Conservative member of Parliament, recently claimed that the policies espoused by the opposition Labour Party would go against the grain of "English exceptionalism". And these are the two components of western culture supposedly most like one another.

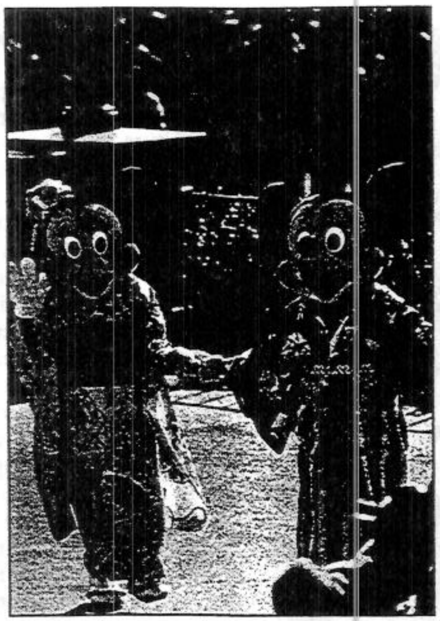
In Islamic countries, the balance between cultural and national identification may be tilted towards the culture. But even here the sense of, say, Egyptian or Iraqi or Palestinian nationhood remains strong. (Consider the competing national feelings unleashed during the Iran-Iraq war.) In other cultures, national loyalty seems pre-eminent: in Mr Huntington's classification, Thailand, Tibet and Mongolia all count as 'Buddhist'. It is hard to imagine that a Thai, a Tibetan and a Mongolian really have that much in common.

So the test of subjective identification is hard to apply. That apart, the writers define a culture in the usual terms: language, religion, history, customs and institutions and so on. Such multiple definitions ring true. As Bernard Lewis's man in the Levantine café suggests, cultures are not singular things: they are bundles of characteristics.

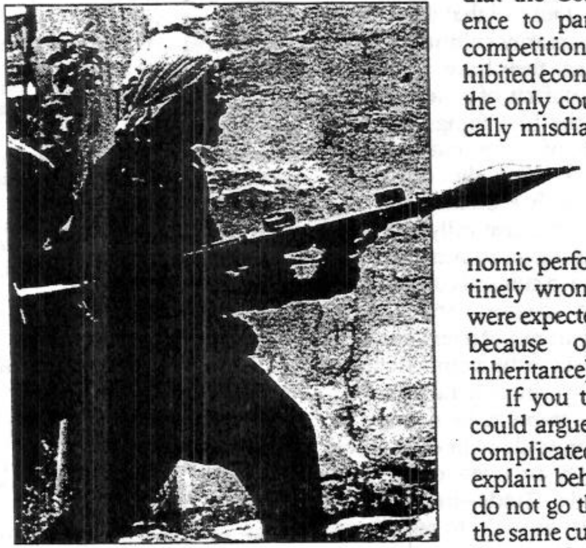
The trouble is that such characteristics are highly ambiguous. Some push one way, some another.

Culture as muddle

Islamic values, for instance, are routinely assumed to be the antithesis of modernising western ones. In Islam, tradition is good; departure from tradition is presumed to be bad until proven otherwise. Yet, at the same time, Islam is also a monotheistic religion which encourages rationalism and science. Some historians have plausibly argued that it was the Islamic universities of medieval Spain that kept science and rationalism alive during Europe's Dark Ages, and that Islam was a vital medieval link between the ancient world of Greece and Rome and the Renaissance. The scientific-rationalist aspect of Islam could well come to the fore again.



Cultured but confused



Not cultured

If you doubt it, consider the case of China and the "Confucian tradition" (a sort of proxy for Asian values). China has been at various times the world's most prosperous country and also one of its poorest. It has had periods of great scientific innovation and times of technological backwardness and isolation. Accounts of the Confucian tradition have tracked this path. Nowadays, what seems important about the tradition is its encouragement of hard work, savings and investment for the future, plus its emphasis on co-operation towards a single end. All these features have been adduced to explain why the tradition has helped Asian growth.

To Max Weber, however, the same tradition seemed entirely different. He argued that the Confucian insistence on obedience to parental authority discouraged competition and innovation and hence inhibited economic success. And China is not the only country to have been systematically misdiagnosed in this way. In countries as varied as Japan, India, Ghana and South Korea, notions of cultural determination of economic performance have been proved routinely wrong (in 1945, India and Ghana were expected to do best of the four—partly because of their supposed cultural inheritance).

If you take an extreme position, you could argue from this that cultures are so complicated that they can never be used to explain behaviour accurately. Even if you do not go that far, the lesson must be that the same culture embraces such conflicting features that it can produce wholly different effects at different times.

That is hard enough for the schools of culture to get to grips with. But there is worse to come. For cultures never operate in isolation. When affecting how people behave, they are always part of a wider mix. That mix includes government policies, personal leadership, technological or economic change and so on. For any one effect, there are always multiple causes. Which raises the second fundamental doubt about cultural explanations: how do you know whether it is culture—and not something else—that has caused some effect? You cannot. The problem of causation seems insoluble. The best you can do is work out whether, within the mix, culture is becoming more or less important.

Culture as passenger

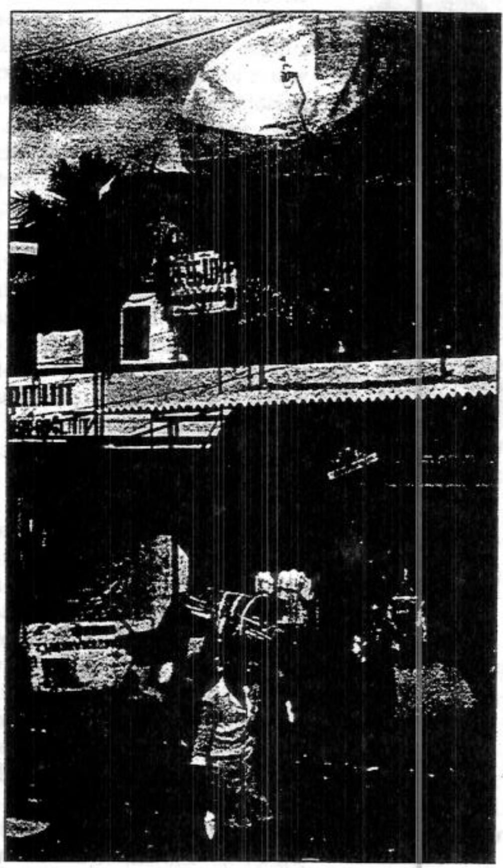
Of the many alternative explanations for events, three stand out: the influence of ideas, of government and what might be called the "knowledge era" (shorthand for globalisation, the growth of service-based industries and so forth). Of these, the influence of ideas as a giant organising principle is clearly not what it was when the cold war divided the world between communists and capitalists. We are all capitalists now. To that extent, it is fair to say that the ideological part of the mix has become somewhat less important—though not, as a few people have suggested, insignificant.

As for the government, it is a central thesis of the cultural writers that its influence is falling while that of culture is rising: cultures are in some ways replacing states. To quote Mr Huntington again "peoples and countries with similar cultures are coming together. Peoples and countries with different cultures are coming apart."

In several respects, that is counter-intuitive. Governments still control what is usually the single most powerful force in any country, the army. And, in all but the poorest places, governments tax and spend a large chunk of GDP—indeed, a larger chunk, in most places, than 50 years ago.

Hardly surprising, then, that governments influence cultures as much as the other way around. To take a couple of examples. Why does South Korea (a low-trust culture, remember) have so many internationally competitive large firms? The answer is that the government decided that it should. Or another case: since 1945 German politicians of every stripe have been insisting that they want to "save Germany from itself"—an attempt to assert political control over cultural identity.

South Korea and Germany are examples of governments acting positively to create something new. But governments can act upon cultures negatively: ie, they can destroy a culture when they collapse. Robert Kaplan, of an American magazine *Atlantic Monthly*, begins his book, "The



Technology's tentacles reach everywhere

Ends of the Earth", in Sierra Leone: "I had assumed that the random crime and social chaos of West Africa were the result of an already-fragile cultural base." Yet by the time he reaches Cambodia at the end of what he calls "a journey at the dawn of the 21st century" he is forced to reconsider that assumption:

> Here I was … in a land where the written script was one thousand two hundred years old, and every surrounding country was in some stage of impressive economic growth. Yet Cambodia was eerily similar to Sierra Leone: with random crime, mosquito-borne disease, a government army that was more like a mob and a countryside that was ungovernable.

His conclusion is that "The effect of culture was more a mystery to me near the end of my planetary journey than at its beginning." He might have gone further: the collapse of governments causes cultural turbulence just as much as cultural turbulence causes the collapse of governments.

Culture as processed data

Then there is the "knowledge era". Here is a powerful and growing phenomenon. The culture writers do not claim anything different. Like the Industrial Revolution before it, the knowledge era—in which the creation, storage and use of knowledge becomes the basic economic activity—is generating huge change. Emphasising as it does rapid, even chaotic, transformation, it is anti-traditional and anti-authoritarian.

Yet the cultural exponents still claim that, even in the knowledge era, culture remains a primary engine of change. They do so for two quite different reasons. Some claim that the new era has the makings of a world culture. There is a universal language, English. There are the beginnings of an international professional class that cuts across cultural and national boundaries: increasingly, bankers, computer programmers, executives, even military officers are said to have as much in common with their opposite numbers in other countries as with their next-door neighbours. As Mr Fukuyama wrote in his more famous book: the "unfolding of modern natural science … guarantees an increasing homogenisation of all human societies." Others doubt that technology and the rest of it are producing a genuinely new world order. To them, all this is just modern western culture.

Either way, the notion that modernity is set on a collision course with culture lies near the heart of several of the culture writers' books. Summing them up is the title of Benjamin Barber's "Jihad versus McWorld". In other words, he argues that the main conflicts now and in future will be between tribal, local "cultural" values (Jihad) and a McWorld of technology and democracy.

It would be pointless to deny that globalisation is causing large changes in every society. It is also clear that such influences act on different cultures differently, enforcing a kind of natural selection between those cultures which rise to the challenge and those which do not.

But it is more doubtful that these powerful forces are primarily cultural or even western. Of course, they have a cultural component: the artefacts of American culture are usually the first things to come along in the wake of a new road, or new television networks. But the disruptive force itself is primarily economic and has been adopted as enthusiastically in Japan, Singapore and China as in America. The world market is not a cultural concept.

Moreover, to suggest that trade, globalisation and the rest of it tend to cause conflict, and then leave the argument there, is not enough. When you boil the argument down, much of it seems to be saying that the more countries trade with each other, the more likely they are to go to war. That seems implausible. Trade—indeed, any sort of link—is just as likely to reduce the potential for violent conflict as to increase it. The same goes for the spread of democracy, another feature which is supposed to encourage civilisations to clash with each other. This might well cause ructions within countries. It might well provoke complaints from dictators about "outside interference". But serious international conflict is a different matter. And if democracy really did spread round the world, it might tend to reduce violence; wealthy democracies, at any rate, are usually reluctant to go to war (though poor or angrily nationalist ones may, as history has shown, be much less reluctant).

In short, the "knowledge era" is spreading economic ideas. And these ideas have three cultural effects, not one. They make cultures rub against each other, causing international friction. They also tie different cultures closer together, which offsets the first effect. And they may well increase tension within a culture-area as some groups accommodate themselves to the new world while others turn their back on it. And all this can be true at the same time because cultures are so varied and ambiguous that they are capable of virtually any transformation.

The conclusion must be that while culture will continue to exercise an important influence on both countries and individuals, it has not suddenly become more important than, say, governments or impersonal economic forces. Nor does it play the all-embracing defining role that ideology played during the cold war. Much of its influence is secondary, ie, it comes about partly as a reaction to the "knowledge era". And within the overall mix of what influences people's behaviour, culture's role may well be declining, rather than rising, squeezed between the greedy expansion of the government on one side, and globalisation on the other.

The books mentioned in this article are:

Benjamin Barber. Jihad versus McWorld (Random House; 1995; 400 pages; $12.95).

Francis Fukuyama. The End of History and the Last Man (Free Press; 1992; 419 pages; $24.95. Hamish Hamilton; £20.) and Trust: The Social Virtues and the Creation of Prosperity (Free Press; 1995; 480 pages; $25. Hamish Hamilton; £25).

Lawrence E. Harrison. Who Prospers? How Cultural Values Shape Economic and Political Success (Basic Books; 1992; $288 pages; $14.).

Samuel Huntington. The Clash of Civilisations? *Foreign Affairs* Vol. 72 (Summer 1993) and The Clash of Civilisations and the Remaking of World Order (Simon & Schuster; 1996; 367 pages; $26).

Robert Kaplan. The Ends of the Earth (Random House; 1996; 475 pages; $27.50. Papermac; £10).

Bernard Lewis. The Middle East (Wiedenfeld & Nicolson; 1995; 433 pages; £20. Simon & Schuster; $29.50).

Seymour Martin Lipset. American Exceptionalism (Norton; 1996; 352 pages; $27.50 and £19.95).

Robert Putnam. Making Democracy Work: Civic Traditions in Modern Italy (Princeton, 1993, 288 pages, $24 and £18.95).

Thomas Sowell. Race and Culture: A World View (Basic Books, 1994, 331 pages, $14).

WORLD COMPETITION

Law and Economics Review

June 1988 Number 33

CONTENTS

ARTICLES

The Private Investor Theory: ITA Equity and Credit Decisions since 1984 — *Richard C. Henderson* 5

The Regulation of Resale Price Maintenance Arrangements in Nigeria — *O. A. Fagbemi* 29

Towards a Single Trans-Tasman Market: A Lawyer's Perspective — *James Farmer* 39

Steel Committee in the OECD: Trade Friction and Multilateral Cooperation — *Teruyoshi Yamazaki* 57

Coping with Extraterritoriality Disputes — *Joseph P. Griffin and Michael R. Calabrese* 75

DOCUMENTS

United Kingdom
Review of Restrictive Trade Practices Policy 97

Japan
The Main Policy of the Fair Trade Commission for 1988 117

Coping With Extraterritoriality Disputes

Joseph P. GRIFFIN[*] and Michael R. CALABRESE[**]

I. INTRODUCTION

Whenever the United States attempts to apply its national law in other sovereign states the action is always controversial and, frequently, is considered offensive or even intolerable by the other states. The extraterritorial application of U.S. export controls has often resulted in a direct conflict between U.S. and foreign laws. Indeed, since the United States began systematically restricting exports in peacetime for national security and foreign policy purposes, there has been foreign opposition to the extraterritorial reach of U.S. export controls.[1]

The U.S. Government's efforts in 1981 and 1982 to delay completion of the Soviet Union's natural gas pipeline to Western Europe raised foreign concerns to new levels. The jurisdictional reach of the pipeline sanctions was in several respects unprecedented, and the foreign response was, accordingly, equally unprecedented. Both the British and French Governments took action to block the application of the U.S. controls, while other European Governments openly urged that U.S. law be defied.

Beyond the political struggles reflected in trade sanctions, and beyond disputes over the proper bounds of state jurisdiction in international law, conflicts of national jurisdiction require painful choices on the part of those engaged in international commerce. The European challenge to the pipeline controls resulted in severe penalties being imposed on affected companies exporting from the United States. Due to strong domestic and international pressure the measures were at first partially abated and ultimately terminated by President Reagan. However, this is hardly a successful solution to a recurring problem for exporters from the United States and their customers. Although the pipeline controls have been withdrawn, and U.S. officials claim to be more sensitive to jurisdictional conflicts, they have not dismantled or repudiated the statutory and regulatory regime used to exert extensive U.S. extraterritorial control over exports.

In Part II of this article the various bases upon which U.S. extraterritorial jurisdiction is asserted will be discussed. Part III illustrates how the breadth of U.S.

This article was first published in the *Journal of World Trade*, Vol. 22 no. 3, June 1988. It is reprinted here by permission of the publisher.

[*] Partner, Morgan, Lewis & Bockius (Washington, D.C.). Chairman of the A.B.A. Section of International Law and Practice.

[**] Associate, Morgan, Lewis & Bockius (Washington, D.C.).

[1] See, e.g., Abbot, *Defining the Extraterritorial Reach of American Export Controls: Congress as Catalyst*, 17 Cornell Int'l L. J. 79, 81 (1984); Berman & Garson, *United States Export Controls—Past, Present and Future*, 67 Colum. L. Rev. 791, 791–92, 867–72 (1967).

WHARTON REPROGRAPHICS

jurisdiction affects U.S. exporting businesses and their customers through a brief review of routine enforcement cases and of the Siberian gas pipeline dispute—the high-water mark of U.S. extraterritoriality. In Part IV foreign governmental reactions to U.S. jurisdictional claims, including so-called "blocking" legislation, judicial decisions and diplomatic efforts are examined in detail. The last section, Part V, draws on the lessons of the past to bring together some practical techniques for counseling the client caught between U.S. extraterritoriality and foreign government defiance.

II. The Bases for U.S. Jurisdiction

The United States primarily relies on the U.S. nationality of the exporter and the U.S. origin of the item or technology exported to justify its extraterritorial control over exports from the United States.[2] This "string of nationality" theory finds support in the legislative history of the Export Administration Act ("the Act").[3] The Department of Commerce ("Commerce" or "the Department") also relies on the pattern of regulatory requirements to which exporters and importers must submit to establish jurisdiction by "acquiescence". The U.S. Government has also argued that a foreign firm's use of an item, previously exported by a person subject to U.S. jurisdiction, is a part of a "continuing export" transaction that remains forever subject to U.S. licensing jurisdiction. None of these theories is widely accepted in international law as a proper basis for extraterritorial jurisdiction over foreign companies and U.S.-origin items wherever they are found. Some of these theories are highly controversial.

There is, however, little doubt that under U.S. law the Department has broad power to impose and enforce export controls beyond U.S. borders. The Department could completely deny the privilege to export if it chose to do so. It also may condition the privilege on the observance of restrictions as to the destinations to which commodities and technologies may be exported and reexported. Unauthorized diversions, even by foreign nationals, may therefore be punished as a violation of U.S. law.[4] Under prior export control statutes, such restrictions were routinely imposed and enforced against domestic and foreign parties.[5] As early as 1970 the Department promulgated regulations prohibiting the reexport to various countries of "any foreign produced direct product of U.S. technical data" if export of the data or product directly would require a validated license.[6]

---

[2] See Note, *Predictability and Comity: Toward Common Principles of Extraterritorial Jurisdiction*, 98 Harv. L. Rev. 1310, 1316–18 (1985).

[3] 50 U.S.C.A. app. §§2401–20 (West Supp. 1987).

[4] See, e.g., Morse & Powers, *U.S. Export Controls and Foreign Entities: The Unanswered Questions of Pipeline Diplomacy*, 23 Va. J. Int'l L. 537, 544–45 (1983); Note, *Extraterritorial Application of the Export Administration Act Under International and American Law*, 81 Mich. L. Rev. 1308, 1311–14 (1983).

[5] See, e.g., 16 Fed. Reg. 315 (1951) (destination control regulations) (superseded). See also *Hydrocarbon Research, Inc.*, 27 Fed. Reg. 12,487 (1962); *Sudexport and General Import Export Co.*, 22 Fed. Reg. 4,512 (1957); *Nison Simon Cohen*, 22 Fed. Reg. 3,134 (1957).

[6] 35 Fed. Reg. 9,184 (1970) (current version at 15 C.F.R. §379.8(a)(3)(1987)).

A. *Nationality*

The Act itself provides no express authority for controls over reexports or other transfers outside the United States, other than general authority over persons, goods, technology and other information "subject to the jurisdiction of the United States". Section 5(a)(1) of the Act authorizes the President to "prohibit or curtail the export of any goods or technology subject to the jurisdiction of the United States or exported by any person subject to the jurisdiction of the United States" for national security purposes.[7] Under section 6(a)(1) of the Act, the President may "prohibit or curtail the exportation of any goods, technology, or other information subject to the jurisdiction of the United States or exported by any person subject to the jurisdiction of the United States, to the extent necessary to further significantly the foreign policy of the United States."[8]

The key phrase "subject to the jurisdiction" of the United States is not defined in the statute. The authority to extend national security and foreign policy controls to the reexport of U.S.-origin goods and technologies from other countries is therefore based on the inference that items subject to U.S. jurisdiction when they leave the United States remain subject to it as they move from country to country in international commerce. The statutory basis for this inference, however, is unclear.

The Act explicitly grants authority only over exports. However, section 5(a)(3) directs that in issuing regulations to carry out national security controls, "particular attention shall be given to . . . the need to take effective measures to prevent the reexport of critical technologies from other countries to countries that pose a threat to the security of the United States".[9] Relying on this provision, on the Act's legislative history, and on long-standing agency practice, the Export Administration Regulations ("Regulations")[10] control reexports of U.S.-origin goods and technology. However, the reference in section 5(a)(3) only to "technologies" and not to "goods" is unexplained and therefore creates some ambiguity as to the section's coverage and its proper interpretation. Moreover, the term "measures" does not necessarily mean unilaterally imposed legal controls and could be read to mean multilaterally negotiated measures.

Commentators agree, however, that Congress intended to exercise broad extraterritorial jurisdiction over reexports of U.S.-origin goods and technology.[11] In 1979, when the renewal of 1977 export control legislation was before Congress, the Senate considered an amendment which would have prohibited controls on non-U.S.

---

[7] 50 U.S.C.A. app. §2404(a)(1) (West Supp. 1987).

[8] *Id.* at §2405(a)(1).

[9] *Id.* at §2404(a)(3).

[10] Codified at 15 C.F.R. pts. 370–399.2 (1987). The Regulations are also published annually in looseleaf form and supplemented by the *Export Administration Bulletin*. Citations herein are from the 1987 looseleaf version.

[11] See, e.g., Marcuss & Mathias, *U.S. Foreign Policy Export Controls: Do They Pass Muster Under International Law?*, 2 Int'l Tax & Bus. Law. 1, 11–14 (1984); Morse & Powers, *supra* note 4, at 544–45; Note, *supra* note 4, at 1311–14.

exports of foreign subsidiaries of U.S. companies. The Senate Report noted the diplomatic friction generated by claims of U.S. jurisdiction over foreign subsidiaries, but the Senate nevertheless rejected the amendment.[12] By doing so, the Senate left intact the language "person subject to the jurisdiction" of the United States. When this language was added by amendment in 1977 it was intended to "control non-U.S.-origin exports by foreign subsidiaries of U.S. concerns".[13] Thus, in 1977 Congress expressed and in 1979 reemphasized its intention to assert export control jurisdiction over subsidiaries of U.S. companies based in foreign nations.

Congress also intended the Act to apply extraterritorially to goods and technology that originate in the United States. The House Report on the 1977 amendments states that the power to control subsidiaries "is in addition to the authority currently provided in the Act for control over the exports of U.S.-origin goods and technology whether from the U.S. or abroad".[14] Read in the light of the history of extraterritorial reexport controls under earlier export control statutes, it seems plain that Congress intended the extraterritorial application of export controls to apply to all goods and technology originating in the United States, wherever they may be found.

Most foreign governments agree that a state may appropriately assert jurisdiction based on the nationality of the exporter. They disagree as to its application to foreign corporations owned or controlled by U.S. persons, which they consider to be nationals not of the United States, but of the country of incorporation or location of the principal administrative office. No foreign government agrees with the U.S. assertion of jurisdiction on the basis of the "nationality" of the goods.[15]

Since the 1982 withdrawal of the pipeline controls, the Regulations have been predicted solely on the U.S. origin of goods or technology. They contain no provisions asserting control over transactions in foreign-origin items on the basis of the U.S. nationality of the exporter. Nevertheless, the statutory framework that gave rise to the pipeline controls still remains in place. The Act still reaches any person and any goods or technology "subject to the jurisdiction of the United States", and that language "is as potent as it was before the sanctions experience".[16] The U.S. Government's construction of the statute has not been repudiated by it or rejected by the courts. It may be argued that Congress, when it enacted the 1985 amendments to the Act, impliedly endorsed the Government's reading of the statute by its reenactment of the jurisdictional language and its tacit acceptance of the Department's widely-publicized administrative interpretation.[17]

---

[12] See S. Rep. No. 169, 96th Cong., 1st Sess. 5 (1979).

[13] S. Rep. No. 466, 95th Cong., 1st Sess. 6 (1977).

[14] H.R. Rep. No. 459, 95th Cong., 1st Sess. 17 (1977).

[15] See pt. IV *infra*.

[16] Ellicott, *Extraterritorial Trade Controls—Law, Policy, and Business*, in *Private Investors Abroad—Problems and Solutions in International Business* in 1983 1, 33 (M. Landwehr ed. 1983).

[17] See 2A Sutherland Statutes and Statutory Construction §49.09 (Sands 4th ed. 1984) ("The United States Supreme Court has held that an administrative interpretation of a provision upon reenactment has the force and effect of law"). But see also Marcuss & Mathias, *supra* note 11, at 15 & 17.

B. *"Submission" to U.S. Jurisdiction*

In addition to controlling goods or technology based on the nationality of a person or the item, the United States also asserts export control jurisdiction based on "acquiescence", or on so-called "submission agreements" by domestic and foreign entities. The Regulations require that "no person in a foreign country" reexport commodities and technical data without U.S. authorization, and foreign importers are required to make explicit commitments not to reexport to unauthorized destinations without the approval of U.S. authorities. These representations generally are required to enable U.S. exporters to obtain individual validated licenses for exports to non-governmental agencies in most countries in Europe, Africa and Asia, non-CoCom (Coordinating Committee for Multilateral Security Export Controls) countries, and to gain approval as consignees under distribution licenses.[18]

U.S. exporters must obtain similar commitments in private contracts with importers of technical data exported under General License GTDR. Further, licensees are required to agree in every application to their strict accountability for use of the license in accordance with the Regulations and the terms and conditions of the license itself.[19] By complying with this pattern of regulations, contractual commitments and written assurances, both domestic and foreign parties have, in the Department's view, agreed to the latter's control over the export transaction and over any subsequent reexport.[20]

C. *"Continuing Export"*

In addition, Commerce has argued that a foreign firm's use of previously exported technical data constitutes a part of a "continuing export" transaction subject to U.S. jurisdiction. Under this theory, once controls are imposed on the basis of an existing license with a person subject to the jurisdiction of the United States, the Department's authority over the transaction continues as long as the license authority is in effect, and includes controls on the foreign-made end-product of U.S. goods or technical data.[21]

D. *Audits and Investigations*

A final area in which the United States asserts broad extraterritorial jurisdiction in export control matters involves regulatory monitoring and enforcement activity outside the borders of the United States. One example is Commerce's desire to be able to verify that companies within a foreign country, seeking to enjoy the status of approved consignees under a distribution license, have a suitable program in place to prevent unauthorized diversions of controlled commodities or technical data. The Department also asserts sweeping powers to conduct wide-ranging investigations of export control compliance procedures when they are informed or have reason to believe that a diversion has taken place.[22]

---

[18] See Export Administration Regulations, pt. 374 (1987).

[19] See *id.* at §§372.1(c) & 379.4(f).

[20] See Feldman, *The Restructuring of National Security Controls Under the 1985 Amendments to the Export Administration Act: Multilateral Diplomacy and the Extraterritorial Application of United States Law*, 21 Stan. J. Int'l L. 235, 269–71 (1985); Marcuss & Mathias, *supra* note 11, at 15–16.

[21] Defendant's Memorandum of Points and Authorities in Opposition to Application for Temporary Restraining Order at 25, *Dresser Indus., Inc. v. Baldridge* (*sic*), 549 F. Supp. 108 (D.D.C. 1982). See also Marcuss & Mathias, *supra* note 11, at 16; Morse & Powers, *supra* note 4, at 562 n.118; Export Administration Regulations, at §§376.12 (goods) & 379.8 (technical data).

In essence, the Department treats goods, technology, or other information "subject to the jurisdiction of the United States" as including all commodities or information exported at any time from the United States. It attributes U.S. nationality to all U.S.-origin goods and technology and foreign-made end products thereof. Commerce has occasionally justified this approach by analogy to the international protection of intellectual property, and by suggesting that recognized international legal principles of jurisdiction might reasonably be extended to apply in the export control context.[23] However, a clear and indisputable basis for such all-encompassing authority is difficult to find in the statute.

III. THE PRACTICAL EXTENT OF U.S. JURISDICTION

Recent enforcement cases such as *in re Datapoint International, GmbH,*[24] *in re Systime Computers Ltd.,*[25] *in re Sven Olof Hakanson,*[26] and *in re Datasaab Contracting, A.B.,*[27] illustrate some of the practical effects of U.S. jurisdictional claims on international exporters and importers. In the *Datapoint* case, the Department charged that the Austrian subsidiary of a U.S. corporation shipped low-level U.S.-origin computer equipment to several East-bloc countries without reexport authorization. The respondent in *Systime* was a British company charged with unauthorized reexports of U.S.-origin goods from the United Kingdom to various controlled destinations, including Syria and Libya. Both firms agreed to settle with the Department, Datapoint agreeing to a civil penalty of $275,000, and Systime agreeing to a $600,000 civil penalty. In addition, Systime was denied all exporting privileges to Country Groups Q, S, W, Y, and Z, generally subject to validated licensing, for a period of two years. In the *Hakanson* case the respondent was charged with unauthorized reexports of U.S.-origin computers from the Federal Republic of Germany through Sweden to the Soviet Union. Hakanson maintained that he had no knowledge of the reexport authorization, but otherwise did not defend the charges. Hakanson's export privileges were completely denied for twenty years.

In *Datasaab*, one of the most notorious of the recent cases, a Swedish firm sought a U.S. license to ship an advanced radar system to the Soviet Union, and shipped approximately $3 million in U.S. parts and equipment in violation of the restrictions imposed by the license even before it was granted. Datasaab raised no jurisdictional objections, pleaded *nolo contendere*, and settled the case by consent decree. The U.S. District Court for the District of Columbia imposed a $3.12 million fine on the company, i.e. the value of the goods exported. Datasaab was also ordered to pay a civil penalty of $2.12 million (the difference between $1 million and the amount of the criminal fine imposed by the District Court). In addition, all of Datasaab's outstanding validated export licenses were revoked and Datasaab was denied export privileges for an indefinite period. The violations were disclosed to U.S. authorities by another Swedish company, Telefonaktiebolaget L M Ericsson, when it acquired Datasaab. L M Ericsson consented to U.S. jurisdiction.[28]

The *Datapoint*, *Systime*, *Hakanson* and *Datasaab* cases all involved wholly foreign conduct by foreign parties. In *Systime*, *Datasaab*, and *Hakanson*, the respondents were not controlled by U.S. persons and had no connection with the United States other than their purchase of U.S. commodities and technologies. In each case there were respectable arguments against the assertion of U.S. enforcement jurisdiction. Yet, in each instance, the respondent chose not to press a jurisdictional challenge and to terminate the case by consent agreement, a decision probably motivated more by economic necessity than by agreement on legal principles.

In December 1981 and June 1982, the United States imposed some of the most far-reaching and controversial U.S. trade sanctions ever employed and several foreign companies found that "giving in" to U.S. jurisdiction was not an option.

The Yamal pipeline dispute has been amply discussed elsewhere and the full chronology will not be repeated in detail here.[29] It is sufficient for present purposes to recall that in retaliation for the Soviet role in the imposition of martial law in Poland during December 1981, the United States imposed two-phased economic sanctions against the Soviet Union. The first set of measures expanded existing oil and gas controls applicable to the Soviet Union to encompass goods and technology for the transportation, as well as the exploration and production, of petroleum and natural gas. Such equipment and technology, which until then could be exported under general license, became subject to a validated license requirement. The Regulations also suspended the processing of all validated license applications for exports to the Soviet Union and all outstanding licenses and authorizations became subject to review and suspension or revocation.[30] In essence, all exports to the Soviet Union for which a validated license was required were halted.

---

[22] See Export Administration Regulations, §§387.8, 387.13(f) & 373.3(n) (1987).

[23] Remarks of Cecil Hunt, Deputy Chief Counsel for Export Admin., U.S. Dep't Comm., reprinted in Expert Advice on Export Controls 1987 (Seminar Documentation) (Legal Studies and Services Ltd. (U.K.) 1987).

[24] (U.S. Dep't Comm., Int'l Trade Admin., Aug. 14, 1987.)

[25] Docket No. 82–602 DWM (U.S. Dep't Comm., Int'l Trade Admin., May 12, 1986).

[26] Docket Nos. 6678–01, 6678–02 (U.S. Dep't Comm., Int'l Trade Admin., Jan. 2, 1988).

[27] 49 Fed. Reg. 19,090 (1984).

[28] *United States v. Datasaab Contracting, A.B.*, Crim. No. 84–00130 (D.D.C. 1984), reported in U.S. Export Weekly (BNA) 905–06 (May 1, 1984).

[29] See generally Atwood, *The Export Administration Act and the Dresser Industries Case*, 15 Law & Pol'y Int'l Bus. 1157 (1983); Butler, *The Extraterritorial Reach of the United States Export Administration Act*, J. Bus. L. 1983 275 (Schmitthoff ed. 1983); Fagelson & Downey, *The Siberian Pipeline Project: Lessons of the Past and Prospects for the Future*, [1983] 7 Oil & Gas L. & Tax. Rev. 145; KinCannon, *The Dresser Case: One Step Too Far*, 5 N.Y.L. Sch. J. Int'l & Comp. L. 191 (1984); Lebedoff & Raievski, *A French Perspective on the United States Ban on the Soviet Gas Pipeline Equipment*, 18 Tex. Int'l L. J. 483 (1983); Lowe, *Public International Law and the Conflict of Laws: The European Response to the United States Export Administration Regulations*, 33 Int'l & Comp. L. Q. 515 (1984); Marcuss, *Soviet Pipeline Sanctions: The President's Authority to Impose Extraterritorial Controls*, 15 Law & Pol'y Int'l Bus. 1163 (1983); Marcuss & Mathias, *supra* note 11; Merciai, *The Euro-Siberian Gas Pipeline Dispute—A Compelling Case for the Adoption of Jurisdictional Codes of Conduct*, 8 Md. J. Int'l L. & Trade 1 (1984); Morse & Powers, *supra* note 4, at 538–44; Moyer & Mabrey, *Export Controls as Instruments of Foreign Policy: The History, Legal Issues, and Policy Lessons of Three Recent Cases*, 15 Law & Pol'y Int'l Bus. 1 (1983); Note, *Dresser Industries: The Failure of Foreign Policy Trade Controls Under the Export Administration Act*, 8 Md. J. Int'l L. & Trade 122 (1984) Note, *Soviet Reaction to the U.S. Pipeline Embargo: The Impact on Future Soviet Economic Relations with the West*, 8 Md. J. Int'l L. & Trade 144 (1984); Note, *supra* note 4; Discussion, *Extraterritorial Application of U.S. Export Controls—The Siberian Pipeline*, 1983 Proc. Am. Soc'y Int'l L. 241 (1985); Comment, *Extraterritorial Application of U.S. Export Laws: Dresser Industries*, [1983] 2 Wis. Int'l L. J. 138.

On June 18, 1982, President Reagan announced that the coverage of the Soviet oil and gas controls would be broadened to include foreign subsidiaries and licensees of U.S. companies. Prior to June 1982, oil and gas controls covered exports and reexports of U.S.-origin equipment or technology and foreign exports of foreign-origin products which incorporated U.S. technology subject to controls when exported from the United States. The new Regulations issued by Commerce on June 22, 1982, restricted foreign subsidiaries of U.S. firms from exporting wholly-foreign-origin equipment and technology. They also prohibited unaffiliated foreign companies from exporting foreign-origin products that were manufactured with technology acquired through licensing agreements with U.S. companies, whether or not the U.S. technology had been subject to controls at the time of export from the United States.[31]

The June 22 controls were designed to prevent the Soviet Union from obtaining indirectly any U.S. equipment or technology for use in constructing the Yamal natural gas pipeline. Specifically, the measures were aimed at preventing European companies from honoring their contracts with the Soviet Union to supply European pipeline equipment that was built with or incorporated U.S.-origin components or technology. The U.S. Government opposed the pipeline, reversing the policy of the previous administration which had encouraged U.S. firms to participate in the project. It was particularly troubling to the U.S. Government that the supply of Siberian natural gas would increase European energy dependence on the Soviet Union, that European payments for natural gas would provide hard currency for the sluggish Soviet economy, and that this benefit to the Soviets would be built in part with U.S.-origin goods and technology. Although the December oil and gas controls were aimed at a number of targets, including the pipeline, the June controls became known as the "pipeline controls" because their primary objective was to impede its construction.[32]

---

[30] 47 Fed. Reg. 141 (1982), amended 47 Fed. Reg. 51,858 (1982). See generally Morse & Powers, *supra* note 4, at 539, Moyer & Mabrey, *supra* note 28, at 60–69.

[31] 47 Fed. Reg. 27,250, 27,251 (1982), revoked 47 Fed. Reg. 51,858 (1982). See generally Morse & Powers, *supra* note 4 at 539–40; Moyer & Mabrey, *supra* note 28, at 69–92.

[32] Morse & Powers, *supra* note 4, at 538–39; Moyer & Mabrey, *supra* note 28, at 70.

Foreign reaction was swift and strong. On August 12, 1982, the Commission of the European Communities transmitted a diplomatic note and legal comments strongly objecting to the U.S. sanctions.[33] Individual European governments also acted. On August 3, pursuant to The Protection of Trading Interests Act 1980, the British Government ordered the four British companies with the largest pipeline contracts—one British-owned company and three subsidiaries of U.S. companies—not to comply with the U.S. Regulations.[34] On August 26, France issued a direct "requisition" order requiring Dresser (France), S.A. to honor pipeline-related contracts with the Soviet Union.[35] The West Germany government asserted its intention to defy the U.S. sanctions, but did not order its companies to ignore them, stating that it had no clear legal basis at its disposal for forbidding adherence to the embargo. Italy declared that its companies would honor their pipeline contracts, but it neither officially forbade its companies to comply with the U.S. sanctions nor specifically directed them to honor their contracts.[36]

The United States, in turn, responded promptly by penalizing those companies violating U.S. law. Acting under section 11 of the Act and section 388.19 of the Regulations, Commerce issued Temporary Denial Orders against six foreign companies. In what may have been a compromise in the face of the severe foreign opposition, Commerce subsequently limited the Denial Orders so that they applied only to U.S.-origin oil and gas goods or technology. Other exports from the United States were permitted. The orders, as modified, effectively prohibited the firms from participating, directly or indirectly, in any transaction involving controlled U.S.-origin oil and gas goods or technology, while revoking all outstanding validated export licenses for such goods or technology connected with those companies. The prohibition extended not only to transactions involving the Soviet Union, but to all transactions.[37]

U.S. enforcement efforts prompted administrative and judicial challenges to both the controls themselves and to U.S. enforcement. For example, Dresser Industries, Inc. and Dresser (France), S.A., petitioned the Department to vacate the Temporary Denial Order prohibiting Dresser (France) from importing U.S. goods, technology, or services relating to the manufacture of equipment for oil and gas projects. The companies also challenged the controls and the penalties implemented by Commerce in Federal district court. In each case they argued that the denial of export privileges without a hearing was a violation of due process, that it was not rationally related to the objective of the Regulations, and that the denial was unlawful because of its extraterritorial and retroactive effects. Commerce denied the Dresser petition,[38] and the District Court refused Dresser's request for a preliminary injunction on the grounds that the company had failed to make the requisite showings of irreparable harm and a likelihood of success on the merits.[39]

---

[33] European Communities: Comments on the U.S. Regulations Concerning Trade with the U.S.S.R., reprinted in 21 I.L.M. 891 (1982).

[34] For the text of the British order, see 21 I.L.M. 851 (1982).

[35] For an account of the French action, see N.Y. Times, Aug. 27, 1982, at D1, col. 4.

[36] Moyer & Mabrey, *supra* note 28, at 82–83.

[37] See 47 Fed. Reg. 44,604 (1982) (*Mannesmann Anlagenbau Aktiengesellschaft*); 47 Fed. Reg. 44,603 (1982) (*AEG Kanis Turbinenfabrik GmbH*); 47 Fed. Reg. 38,169 (1982), modified 47 Fed. Reg. 42,392 (1982) (*Creusot-Loire, S.A.*); 47 Fed. Reg. 40,205 (1982) (John Brown Eng'g Ltd.); 47 Fed. Reg. 37,709 (1982) (*Nuovo Pignone S.p.A. Industrie Meccaniche Fonderia*); 47 Fed. Reg. 38,170 (1982), amended 47 Fed. Reg. 39,708 (1982) (Dresser (France), S.A.).

In the end, the confrontation was resolved when President Reagan, citing the apparent willingness of the European allies to move in the direction of a unified economic policy toward the Soviet Union, announced on November 13, 1982, the withdrawal of the December 1981 and June 1982 Regulations.[40] This disposition left unresolved the variety of important questions concerning the extraterritorial reach of U.S. export controls.

IV. Foreign Governmental Reactions

Foreign governments resist U.S. extraterritorial jurisdiction in export control matters on policy as well as on strictly legal grounds. Foreign governments, even allied ones, often disagree with the United States about the foreign policy objectives sought to be achieved by sanctions. Many commentators dispute the efficacy of trade sanctions as foreign policy tools and as effective guardians of national security, viewing them as little more than ineffective symbolism.[41] From most of our trading partners' point of view, however, the principal ground for objection is that the practice violates well-established doctrines of international law and impermissibly infringes on the sovereignty of other states. From the U.S. perspective, although international law is unquestionably part of U.S. law, it "bends to the will of Congress" and must give way when it conflicts with a Federal statute.[42] In the case of export controls, Congress clearly intended extraterritorial application, and was aware of the likely response of the international community to such action when it enacted the legislation.[43]

Despite the complete divergence in the U.S. and foreign governments' perspectives, the international legal status of U.S. extraterritorial export controls is important for several reasons. Disputes over extraterritoriality often arise in fora outside the United States, in jurisdictions where international law may be dispositive or

---

[38] 47 Fed. Reg. 51,463 (1982).

[39] *Dresser Indus., Inc. v. Baldridge (sic)*, 459 F. Supp. 108 (D.D.C. 1982).

[40] Moyer & Mabrey, *supra* note 28, at 83–84.

[41] See, e.g., Abbott, *Linking Trade to Political Goals: Foreign Policy Export Controls in the 1970s and 1980s*, 65 Minn. L. Rev. 739, 744 (1981).

[42] See Restatement of Foreign Relations Law of the United States (Revised) §§134, 135 (Tent. Draft No. 6 1985); *The "Over the Top"*, 5 F.2d 838, 842–43 (D. Conn. 1925).

[43] Post-Hearing Brief of U.S. Department of Commerce, *in re Dresser (France) S.A.* at 42, Docket No. 632 (U.S. Dep't Comm., Int'l Trade Admin.).

persuasive when the authority of U.S. law is challenged.[44] Principles of international law can also affect the way that U.S. export policy is formulated at the outset and applied in particular instances. For example, as a result of the pipeline dispute, the 1985 amendments to the Act contained a "contract sanctity" provision that bars, except in extraordinary circumstances, the application of new foreign policy-based controls to exports under contracts entered into or export licenses issued prior to the notice of intention to impose controls.[45] The Department has also expanded permissive reexport authorization for shipments into and among CoCom (Coordinating Committee) countries which helps remove the "extraterritorial stigma" of U.S. controls.[46]

A. *Legal Theories; Foreign Legislative and Judicial Responses*

Neither states nor scholars agree on the requirements of international law applicable to the exercise of extraterritorial jurisdiction.[47] Two bases of state jurisdiction—territoriality and nationality—are commonly accepted, but there is disagreement about their scope. Three other doctrines—the "protective", "passive personality", and "universality" principles—may justify a state's assertion of jurisdiction in circumstances where the territoriality and nationality principles are not applicable.[48]

The passive personality principle asserts that a state has jurisdiction in certain cases to punish aliens for acts committed abroad that are harmful to nationals of that state. The universality principle holds that there is at least one crime—piracy—and there may be others, which justifies punishment anywhere as a matter of international public policy.[49] Neither of these is relevant to the extraterritorial application of export control laws. Thus, if there is a basis for the U.S. position it must be found either in the territorial, nationality or protective principles.

1. *Territoriality*

The territoriality principle provides that a nation has jurisdiction to prescribe rules over persons and acts taking place in its territory.[50] The "objective" territorial principle, is more controversial. While it is widely accepted that a state has jurisdiction over acts or omissions taking place outside its territory when those acts or omissions have effects within its territory and certain conditions are met, there does not seem to be a consensus as to what constitutes those conditions. Mere incidental or minor effects within a state's territory are probably insufficient to justify jurisdiction on the basis of territoriality principles.[51]

---

[44] For example, in Great Britain, international law has no force as domestic law until it is specifically incorporated into the laws of Britain by an Act of Parliament. See I. Brownlie, Principles of Public International Law 45–53 (3d ed. 1979). Italy is an example of a state where a constitutional mandate requires that domestic law confirm to principles of international law. *Id.*

[45] Feldman, *supra* note 19, at 252–53.

[46] See Export Administration Regulations, §374.2 (1987); Remarks of Dr. Paul Freedenberg, Ass't Sec'y for Trade Admin., U.S. Dep't Comm., reprinted in *Expert Advice on Export Controls 1987* (Seminar Documentation) (Legal Studies and Services Ltd. (U.K.) 1987).

[47] See generally A. Herman, *Conflicts of National Laws with International Business Activity: Issues of Extraterritoriality* (1982); D. Rosenthal & W. Knighton, National Laws and International Commerce: The Problem of Extraterritoriality (1982); Extraterritorial Application of Laws and Response Thereto (C. Olmstead ed. 1984); The Extraterritorial Application of National Laws (D. Lange & G. Born eds. 1987); Extraterritorial Jurisdiction: An Annotated Collection of Legal Materials (A. Lowe ed. 1983).

[48] Marcuss & Mathias, *supra* note 11, at 17–18.

[49] *Id.* at 18–19.

More controversial still is the "effects doctrine", which has been widely used by U.S. courts as a basis for jurisdiction in cases involving antitrust law, criminal law, and securities regulation.[52] Although jurisdiction under this principle has some basis in international law, the U.S. position has been subject to criticism because of its potentially limitless scope. The Draft Revised Restatement of Foreign Relations Law limits the doctrine to cases in which, *inter alia*, the effect is direct, foreseeable, and substantial, the rule applied will not be inconsistent with the traditions of the international legal system, and the assertion of jurisdiction is "reasonable".[53] In most situations to which export controls might be applied, there is simply no tangible effect within the United States. Instead, any effect is most likely to be on the United States' perceptions of its national security or foreign policy interests.

Many states reject entirely the effects doctrine as a basis for control over exports, arguing that such extraterritorial action is always illegal under international law. In the British view, for example, a state which regulates or punishes the conduct of foreigners abroad contravenes international law because it infringes upon the sovereignty of the state where the conduct is committed. Moreover, if the actor is not a national of the state where the conduct is committed, the sovereignty of the country of which the actor is a national is infringed upon as well.[54]

The British Government has relied on the Protection of Trading Interests Act 1980 to block or to frustrate various types of foreign extraterritorial interference with British trading interests.[55] When the four British companies affected by the pipeline sanctions were ordered, pursuant to that Act, not to comply with the June U.S. Regulations, they were faced with the dilemma of whether to obey U.S. law, in which case they might be prosecuted and fined under British law, or to obey British law, and risk being punished under U.S. law. Choosing to obey British law, the law of their resident jurisdiction, immediately resulted in Commerce issuing a Temporary Denial Order.

---

[50] I. Brownlie, *supra* note 43, at 287.

[51] Marcuss & Mathias, *supra* note 11, at 18.

[52] See, e.g., *United States v. Aluminum Co. of Am.*, 148 F.2d 416 (2d Cir. 1945) (antitrust); *United States v. Baker*, 609 F.2d 134 (5th Cir. 1980) (criminal law); *Leasco Data Processing Equip. Co. v. Maxwell*, 468 F.2d 1326 (2d Cir. 1972) (securities).

[53] Restatement of Foreign Relations Law of the United States (Revised) §403 (Tent. Draft No. 7, 1986).

[54] See U.K. Sec'y of State for Trade, American Measures Against the Soviet Union: Protection of Trading Interests Act (August 2, 1982), reprinted in 21 I.L.M. 851 (reprinting the Protection of Trading Interests (U.S. Reexport Control) Order 1982).

[55] *Id*; April, *Blocking Statutes as a Response to the Extra-territorial Application of Laws*, in Extra-territorial Application of Laws and Responses Thereto 223 (C. Olmstead ed. 1984).

2. *Nationality*

Most extraterritorial action taken under the Act is based on the nationality principle which grants a state jurisdiction to govern the conduct of its nationals, even when its nationals are outside the state's territory. However, the international rules governing the nationality of corporations are not settled, and there are virtually no rules of international law governing the nationality of goods or technology. This principle therefore furnishes weak support for the extraterritorial extension of export controls.[56]

Corporations

The traditional common law rule attributes to a corporation the nationality of the state of incorporation, while the rule in civil jurisdictions determines the nationality of a company according to its "seat" (*siège social*), i.e., the principal place of business or the location of the company's central administrative offices.[57] The most important pronouncement by the International Court of Justice on the nationality of a corporation for purposes of international law states that a corporation is a national of the state under the laws of which it was incorporated and in whose territory it has its registered office.[58]

According to either of the accepted tests of a corporation's nationality, the U.S. claim of jurisdiction over foreign subsidiaries or foreign corporations is invalid. These companies are neither incorporated under the laws of the United States nor do they have their "seat" in the United States. Thus, they are not U.S. nationals. Even though Congress intended the President to have the power under the Act to control the exports of foreign subsidiaries, this assertion of jurisdiction cannot be justified by the nationality principle.

During the pipeline controversy, the French government did not have legislation similar to The Protection of Trading Interests Act 1980 to deal with extraterritorial interference with French export trade. However, there was some legal precedent for French judicial action to prevent a company of French nationality from breaking an export contract in order to comply with extraterritorial U.S. legislation. In 1964 the French subsidiary of Freuhauf Corporation contracted to sell goods which were ultimately intended for export to China and the U.S. Treasury Department ordered the parent company, a U.S. corporation, to cause the French company to cancel its contract. In *Freuhauf Corp. v. Massardy*,[59] an appeal to the Paris *Cour d'appel* by the

---

[56] Morse & Powers, *supra* note 4, at 561–62; Note, *supra* note 4, at 1321–25.

[57] W. Bishop, *International Law Cases and Materials* 488 (3d ed. 1971); W. Friedman, O. Lissitzyn & R. Pugh, *Cases and Materials on International Law* 513 (1969).

[58] *Barcelona Traction, Light and Power Co., Ltd.* (Belg. v. Spain), 1970 I.C.J. 3 (Judgment of Feb. 5).

[59] Judgment of May 22, 1965, *Cour d'appel*, Paris, reprinted in 5 I.L.M. 476 (1966).

U.S. directors, the *Cour d'appel* approved the French directors' action to block the U.S. parent's cancellation of the contract.[60]

In 1982, the French Government did not wait for Dresser (France) to follow the *Freuhauf* precedent. Its requisition order to Dresser (France) required Dresser (France) to honor its contractual obligations to the Soviet Union or face criminal penalties.[61] Dresser (France) did so, and it, too, became subject to a Temporary Denial Order.

Nationality was judicially rejected as a basis for U.S. jurisdiction over a foreign subsidiary in the Dutch *Sensor*[62] case. Sensor, which was wholly-owned by a U.S. company, contracted to sell gas exploration equipment to a French company for ultimate sale to the Soviet Union. When Sensor informed its buyer that it could not meet its delivery obligation because, as a subsidiary of a U.S. corporation, it had to respect the June regulations, the French firm sought a court order compelling delivery. The District Court in The Hague considered and rejected the proposition that a foreign person owned or controlled by a U.S. person is a person subject to U.S. jurisdiction. The court held that Sensor "has Netherlands nationality, having been organized in the Netherlands under Netherlands law and both its registered office and its real center of administration being located within the Netherlands". Applying Netherlands law, the court ordered performance of the contract.[63]

Goods

There are no rules of international law governing the nationality of goods other than airplanes, ships, and possibly historical cultural artifacts. Attributing any nationality to inanimate objects poses unique difficulties. International law has developed the concept of nationality almost exclusively in the context of individual persons, using principles inapplicable to commercial goods and technologies.[64] On what principled basis can one determine, for example, the "nationality" of pipeline equipment assembled in the United States from components imported from other states?

The application of U.S. export control laws based on a claim that its exports are unalterably of U.S. origin has been rejected by the European Commission[65] and foreign courts. For example, in the *American President Lines*[66] case, the Supreme Court of Hong Kong refused to recognize the right of the United States to prohibit delivery of U.S.-origin goods situated in Hong Kong. Instead, the court found that after their discharge the goods were no longer subject to U.S. jurisdiction, but were governed by the *lex situs*, in this case, the law of Hong Kong. To hold otherwise "would be to create an incursion into the sovereign rights of Hong Kong, which . . . could never have been the intention of the draftsman".[67]

3. *Protective Jurisdiction*

Under the protective principle a state has jurisdiction with respect to crimes committed by aliens outside its territory against the security, territorial integrity, or political independence of that state. A state can prescribe rules of law for conduct outside its territory that threatens its security as a state, provided that the conduct is generally recognized as a crime under the law of states that have reasonably developed legal systems. The Revised Restatement enumerates espionage, counterfeiting of the state's seal or currency, falsification of official documents, perjury before consular officials, and conspiracies to violate the immigration or customs laws as examples of crimes to which the protective principle might be applied.[68] U.S. law has developed other stringent prerequisites to assertion of jurisdiction based on this. First, the conduct in question must generally be recognized as criminal by civilized nationals. Second, "[a]ll cases which have invoked the protective principle . . . [involved] actions where there was a demonstrable effect on the United States *in particular*".[69]

Because of the strict requirements applicable to its use, the protective principle can justify extraterritorial export regulations only in unusual situations. The exports controlled by the Act will rarely be generally considered criminal activity by the international community, and such exports most likely will not impair particular U.S. interests unless the relationship between the United States and the importing national is particularly antagonistic. For example, the pipeline controls were promulgated under the Act's provisions for foreign policy controls which makes it difficult for the United States to justify its jurisdiction on grounds of national security. Moreover, because the controlled exports were neither generally considered criminal by the international community, nor likely to produce effects on the United States in particular, the protective principle cannot justify extraterritorial controls in this situation.

4. *Submission to National Jurisdiction*

The related argument that the exporters and their customers have voluntarily submitted to U.S. jurisdiction is not persuasive as a matter of international law. First, it is open to question how "voluntary" such a submission is, especially when the alternative is the denial of a license and a lost business opportunity. In such a situation it could be reasonably argued that these are more elements of compulsion or duress in the submission than they are of free and full agreement. Second, the fact that an exporter and an importer agree that they will not reexport a product in violation of the regulations is not submission to U.S. jurisdiction wherever they or the goods may be. The United States may have a cause of action against the parties, and may be able to enforce its rights if the goods or parties are present in the United States or if the United States can gain the cooperation of foreign states but, as the European Commission pointed out during the pipeline controversy, this does not mean that the United States has authority under international law to enforce the Regulations in the territory of another state.[70]

---

[60] *Id.* See generally Craig, *Application of the Trading With the Enemy Act to Foreign Corporations Owned by Americans: Reflections on Freuhauf v. Massardy*, 83 Harv. L. Rev. 579 (1970).

[61] See note 34 *supra*.

[62] Judgment of September 17, 1982, *Compagnie Européenne Des Pétroles S.A. v. Sensor Nederland B.V.*, (Dist. Ct., The Hague), reprinted in 22 I.L.M. 66 (1983).

[63] 22 I.L.M. at 71, 74.

[64] Note, *supra* note 4, at 1324 (citing *Nottebohm Case* (Liechtenstein v. Guatemala), 1955 I.C.J. 4).

[65] See note 37 *supra*.

[66] *American President Lines, Ltd. v. China Mut. Trading Co.*, 37 Hong Kong L. Rep. 38 (Sup. Ct. (orig.)), appeal dismissed, 37 Hong Kong L. Rep. 111 (Sup. Ct. (app.) 1953).

[67] 37 Hong Kong L. Rep. 38 (Sup. Ct. (orig.) (1953)). See also *Moens v. Ahlers Am. Lloyd*, 30 R.W. 360 (Trib. Comm., Antwerp 1965).

[68] Revised Restatement, *supra* note 52, at §402, comment (f).

[69] *United States v. James-Robinson*, 515 F. Supp. 1340, 1345 (S.D. Fla. 1981) (emphasis in original).

Third, there is good reason to doubt the authority of a private party voluntarily to submit property that is located in another state to the jurisdiction of the United States. In U.S. jurisprudence, "parties cannot waive lack of [subject matter] jurisdiction by express consent, by conduct, or even by estoppel".[71] The basis for this rule, i.e. that jurisdictional concerns are too fundamental to be left to the litigants, applies even more forcefully in the international area, where jurisdictional issues involve claims of competing sovereigns to dominion and control over items within their territory.

All these objections were presented during the pipeline controversy, when all the affected governments in the European Commission united in the view that the extraterritorial reach of the June 1982 Regulations was contrary to international law.[72] Even those which accepted that a company may be a national of a country other than the country of incorporation did not accept that every company which is owned or controlled by U.S. persons is a national of the United States by reason of that fact alone. Those which accepted the extraterritoriality can be justified in certain cases of national emergency, or under the effects doctrine, did not accept that states have the right to legislate extraterritorially in order merely to advance foreign policy objectives. None accepted the U.S. view that goods and technical data originating in the United States remain subject to U.S. law after export.[73]

B. *Diplomatic Efforts*

Foreign governments angered by the assertion of U.S. extraterritorial jurisdiction have also used diplomatic means to press their views upon the U.S. Government. In addition to the diplomatic protests occasioned by the pipeline sanctions, the European Commission also submitted an *aide-mémoire* to the U.S. Government when 1985 amendments were enacted and has commented on recent proposals to expand certain general licenses.[74] In each case the Commission maintained its objection to provisions that allow extraterritorial enforcement. As a result of these and other protests, the U.S. Government has become increasingly sensitive to the concerns raised by foreign states over extraterritoriality and has begun publicly to commit itself to a process of managing these conflicts of jurisdiction through diplomatic negotiations.[75]

Mutual accommodation through negotiations has several advantages over unilateral remedies. Negotiated limits on U.S. jurisdiction would reduce the level of trade conflicts because they would reflect the essential interests of the major affected states and their companies. Negotiations also allow the prospect of foreign concessions that would not be forthcoming with unilateral solutions. A demonstrated willingness to negotiate might enhance U.S. relations with nations antagonized by the extraterritorial application of U.S. law. The Secretary of State, Commerce Department export control officials and other commentators have endorsed negotiated solutions to extraterritorial conflicts.[76] The CoCom forum seems particularly suited for such talks.

An example of a negotiated solution occurred in 1957 when the U.S. Government demanded that the Ford Motor Company halt its Canadian subsidiaries' participation in the export of automotive parts from Canada to China. The ensuing outcry in the Canadian Parliament against the U.S. effort to regulate Canadian subsidiaries led to an apparent agreement calling for the United States to consult with Canada on a case-by-case basis on licenses to be granted for trade with China.[77]

It is fair to say that the conflicts occasioned by U.S. extraterritorial export controls in recent years have abated somewhat. The United States has avoided extremes of extraterritorial reexport controls in recent foreign policy-based controls imposed on, for example, Libya and Nicaragua.[78] Foreign countries have stepped up their own enforcement efforts in cooperation with U.S. officials, as shown by the recent Toshiba/Kongsberg episode.[79] The United States also appears to be more willing to rely on agreements within CoCom to protect against diversion. However, the U.S. position on extraterritoriality has not been repudiated, nor has it generally been explained or justified. It is clear that extraterritorial export controls create political costs for the nation imposing them and for affected third countries, and impose extreme economic costs on businesses caught in the middle. In the case of the pipeline regulations, foreign governments and foreign courts rejected U.S.

---

[70] See note 32 *supra*.

[71] C. Wright, A. Miller & E. Cooper, Federal Practice and Procedure §3522, at 46–47 (1975).

[72] See note 32 *supra*.

[73] Fagelson & Downey, *supra* note 28, at 147.

[74] *Aide-Mémoire* of the European Community and its Member States Concerning the Renewal of the Export Administration Act of 1979 (April 28, 1983); Delegation of the Commission of the European Communities: Comments on Proposed Amendments to the Export Administration Regulations (August 18, 1986).

[75] See, e.g., Address by Sec'y of State Shultz before the S. C. Bar Ass'n (May 5, 1984), reprinted in Dep't St. Bull. 33–37 (June 1984).

[76] See, e.g., Abbott, *supra* note 1, at 119; Feldman, *supra* note 19, at 277; Address by Sec'y of State Shultz, *supra* note 73, at 36; Remarks of Cecil Hunt, *supra* note 22.

[77] See *Joint Statement on Export Policies*, 39 Dep't St. Bull 209 (1958).

[78] Remarks of Cecil Hunt, *supra* note 23.

[79] See Wendt, *Export Control Policy and CoCom* (Current Pol'y No. 1031, U.S. Dep't of State 1987).

jurisdictional claims on the grounds that those claims were improper under accepted international legal principles. They will presumably continue to do so and thus impair the effectiveness of U.S. export control policy insofar as it is conducted through extraterritorial means.

V. Coping with Extraterritoriality Disputes

The first priority of a U.S. lawyer counseling a foreign company engaged in U.S. export trade must be to determine both the risk and extent of the client's liability for the unlawful diversion U.S.-origin goods or technology. The Act provides sanctions for the violation of any of its provisions, the Regulations and any orders or licenses issued pursuant to the Act. Knowing and wilful violations are punishable, as are evasion, conspiracy and attempts to violate. There are broad provisions covering accessories, and sanctions for misrepresentation and concealment, including an important provision on the continuing effect of representations.[80]

Commerce may impose a civil penalty of up to $10,000 for each violation, unless the violation involves national security controls, for which a civil penalty of up to $100,000 may be imposed. Criminal penalties for knowing violations include a fine of $50,000 or five times the export value, whichever is greater, or imprisonment for up to 5 years, or both. For wilful violations committed by an individual, a fine of up to $50,000 or imprisonment for up to ten years, or both, may be imposed. Wilful violations by other persons are punishable by a fine of $1,000,000 or up to 5 times the export value, whichever is greater.[81]

Commerce also has an arsenal of administrative sanctions, the most powerful of which is the Temporary Denial Order which was the key enforcement measure in the pipeline affair.[82] Such orders are used principally against foreigners, especially where Commerce is unable to get *in personam* jurisdiction over a respondent to impose other penalties effectively. Temporary Denial Orders represent a real threat to foreign companies and, as discussed above, raise serious questions about the proper reach of Commerce's enforcement jurisdiction. However, with the exception of the pipeline case, very few respondents have contested Temporary Denial Orders in legal proceedings. Counsel should try to negotiate the scope of any Temporary Denial Order, an approach that was apparently successful in the pipeline case, and may have played a role in *Datasaab*, as well.

Planning is essential to assure compliance with U.S. law. There are special considerations involved in the administration of an effective export control program for foreign subsidiaries and affiliates. The distances involved alone make monitoring difficult, so close coordination with management is necessary. Also, foreign

---

[80] 50 U.S.C.A. app. §§2410(a)–(c) (West Supp. 1987); Export Administration Regulations, pt. 387 (1987).

[81] 50 U.S.C.A. app. §2410(b) (West Supp. 1987).

[82] *Id.* at §2410(c).

corporate officials may not be particularly familiar with or sensitive to U.S. export controls, so effective education is also a key element. The reexport provisions make up only two substantive parts of the Regulations, but they can raise complex legal, political and administrative issues. Many companies run afoul of U.S. law because the reexport control requirements were not understood clearly or taken seriously, especially with respect to controls on foreign-made products of U.S. technical data.

Counsel also must anticipate that the host country may compel the client by a governmental directive, law, or other mandate to sell, resell, export or reexport U.S.-origin goods or technology, either within or without the host country, in violation of U.S. law. The firm will, in such a case, be faced with a decision whether to violate U.S. law or that of the host country jurisdiction. The range and extent of sanctions that may be imposed in the host country jurisdiction therefore become an important factor in the economic and legal equation. Counsel must also be aware of the potential personal liability of directors, officers and employees of affected companies in both jurisdictions.[83] The client should be advised well in advance of potential conflicts to consider all of these issues and to develop corporate policies to be followed in case of such a conflict.

In negotiating an export contract it may be advisable to structure the transaction using back-to-back contracts, the first with the U.S. supplier of goods or technology and the second with the importing foreign company. The terms and conditions of each contract should "dovetail" and, in particular, define *force majeure* to include changes in the application of export control laws and foreign decrees or other mandates that may require performance of the contract.

No amount of preventive planning can eliminate the risk that a host government will place the foreign client in the untenable position of being subject to two incompatible legal rules. In the pipeline controversy, the sovereign compulsion doctrine was raised as a possible defense, but its applicability was not settled. The doctrine arguably precludes the United States from enforcing laws or regulations against persons who commit acts outside the United States that are required by the law of the state where those acts are committed.[84] This principle has been recognized by U.S. courts in fashioning decrees against foreign entities to accommodate the legitimate concerns of their sovereigns.[85]

Several facts made the applicability of the sovereign compulsion doctrine to the pipeline cases problematic. There were contentions that some European firms had procured blocking orders from their governments. This action, if true, would have undercut the equitable argument for recognition of the sovereign compulsion defense. Whether the sovereign compulsion doctrine applies in administrative proceedings is also uncertain. Thus, it is far from clear that an administrative refusal to accept a sovereign compulsion justification for violation of the pipeline regulations would have been an error.[86]

---

[83] See Stebbings, *Export Controls: Extraterritorial Conflict—The Dilemma of the Host Country Employee*, 19 Case W. Res. J. Int'l L. 303 (1987).

[84] See, e.g., *Timberlane Lumber Co. v. Bank of Am. N.T. & S.A.*, 549 F.2d 597, 606 (9th Cir. 1976); *Interamerican Refining Corp. v. Texaco Maracaibo, Inc.*, 307 F. Supp. 1291, 1298 (D. Del. 1970); *United States v. General Elec Co.*, 115 F. Supp. 835, 878 (D.N.J. 1953); Griffin, *Zenith Leaves Major International Antitrust Questions Unanswered*, Int'l. Bus. Law. 354–56 (November 1986); Motion for Leave to File Brief *Amici Curiae* and Brief of the Governments of Australia, Canada, France, and The United Kingdom of Great Britain and Northern Ireland in *Matsushita Elec. Indus. Co., Ltd. v. Zenith Radio Corp.*, 475 U.S. 574 (1986).

[85] *United States v. Watchmakers of Switz. Information Center, Inc.*, 1963 Trade Cas. (CCH) ¶7,600 (S.D.N.Y. 1962), modified 1965 Trade Cas. (CCH) ¶71,352 (S.D.N.Y. 1965).

When faced with irreconcilable legal obligations, companies should be prepared to fight a somewhat inconsistent legal battle, arguing first in the foreign jurisdiction that local law should not be imposed because of U.S. jurisdiction, and then, after losing in the host jurisdiction, arguing that the assertion of U.S. jurisdiction over a foreign corporation is improper.

The international legal theories discussed above have in the past provided a measure of relief in foreign courts to foreign subsidiaries caught between inconsistent legal rules. In turn, the U.S. parent of a sanctioned foreign company may argue that principles of comity require the United States to withdraw its assertion of jurisdiction. The international comity principle, as codified in the Revised Restatement, provides that a state may not apply its laws to the conduct of persons having connection with another state when the exercise of such jurisdiction is unreasonable, based on an assessment of the interests of and connections between those sovereigns and the conduct in question.[87] In application, this principle means that a state that has a colorable basis for regulating conduct within the territory of another state nevertheless must abstain from extraterritorial regulation if other states have a closer connection to the transaction in question.

In addition to these strictly legal approaches, political or diplomatic options should be pursued. The pipeline controversy itself was resolved by a combination of diplomacy and domestic political reaction brought to bear on the U.S. Government. Foreign companies should maintain a close liaison with their government and embassy in Washington. They should not assume, however, that the invocation of foreign governmental interest will resolve the matter. Most governments expect their business to participate actively in the protection of their own interests. The type of foreign government involvement should be tailored to reflect whether the case is one of the more "routine" ones, or is a highly-charged "political" case. "Quiet diplomacy" may be more successful than a frontal attack in resolving "routine" disputes.

VI. CONCLUSION

Although the extraterritorial reach of U.S. export controls is controversial, the U.S. Government probably will continue to assert its authority to regulate export transactions beyond the borders of the United States. The arguments advanced in the pipeline case that such extraterritorial application violates international law may excuse foreign companies from compliance with U.S. controls in foreign jurisdictions, but probably will not be useful in an enforcement action brought by the U.S. Government. The best hope for a resolution of the problems arising out of jurisdictional conflicts lies in bilateral or multilateral negotiations, perhaps under CoCom auspices.

---

[86] Morse & Powers, *supra* note 44, at 564–65 n.124.

[87] Revised Restatement, *supra* note 52, at §403.

85
107
WHARTON REPROGRAPHICS 7

LIPPINCOTT LIBRARY
DOCUMENT DELIVERY SERVICE

We've all heard stories about U.S. exporters who shipped goods, then were not paid—or importers who bought goods from abroad, then found them not up to quality or received them too late. These problems are often due to too much concentration on the glamorous side of international trade and too little attention to the practical.

The Practical Side of International Trade

BY THOMAS B. McVEY

For an international trade transaction to be successful, the financial side of the transaction must be well structured. Financing has two parts: the operational aspects of making or receiving payments for products sold in foreign markets, and the credit arrangements for overseas buyers to use in purchasing the products being sold. If these two aspects are properly handled, even the most complex transaction can be conducted on a safe and profitable basis.

In this article, we will describe the operational aspects of trade finance for exporters and importers, including the technical aspects of arranging for payments in trade transactions. In the next issue of this magazine, we will focus on the arrangement of supplier credits to foreign buyers to facilitate overseas sales, including a description of financing available through government agencies such as the U.S. Export-Import Bank, private sources of financing, a'forfait financing techniques, the use of export subsidies, and "alternative" forms of trade finance such as countertrade, barter, and "blocked funds" repatriation. The techniques described will apply to an export transaction, although the same methods can be applied in reverse in an import transaction.

Unique Nature of International Transactions

International business transactions present special business, financial, and legal risks not present in domestic transactions. Because business is conducted in two or more sovereign countries, the legal system of the first country may not adequately protect the right of parties located in the second country and vice versa. Parties cannot always rely on conventional legal remedies, such as litigation or attachment of assets.

Because of this, international business has developed over the years special procedures to safeguard transactions against risks such as nonpayment, late delivery, delivery of defective goods, and foreign exchange risk. These are based on a combination of business procedures, financial instruments, and legal principles that are tightly interwoven through custom and international convention rather than derived from the legal system of any one nation. The methods of payment described below are the cornerstone of these practical procedures and around which all other aspects of international trading transactions revolve.

Methods of Payment

The following are the principal methods of payment used in trade transactions.

Cash in advance. The safest method available for the exporter to receive payment for his overseas sale is payment in advance. Here, the foreign purchaser pays the full purchase price to the exporter prior to shipment of the goods. This is great for the exporter, but it is extremely risky for the overseas buyer. In addition, it permits competing sellers to offer more favorable payment terms, thereby allowing them to be more competitive in selling to the overseas buyer. For these reasons, payment in advance is not as regularly used as other forms of payment.

Letters of credit. This is perhaps the most common form of payment in trade transactions. Under this method, the foreign buyer (the "account party") asks its bank to issue a letter of credit to the U.S. exporter (the "beneficiary"). The letter states, in effect, that the seller will ship the requested products and forward certain shipping documents to the foreign bank, and the bank in turn will pay the amount in the letter to the exporter. This provides a form of assurance to the seller that once it ships its products, a creditworthy financial institution will stand behind payment for the goods. Because of the reliance on the transfer of documents to trigger the payment obligation, this type of letter of credit is referred to as "commercial" or "documentary" letters of credit.

Of course, the letter of credit is only as valuable as the financial condition of the bank issuing it. Consequently, exporters must be sure that the issuing bank is reputable and creditworthy in world financial markets. One method of assuring this is to have the letter of credit "confirmed" by a U.S. bank. In this case, the foreign bank will send its letter of credit to the exporter's bank in the United States, and the U.S. bank will issue its own commitment to pay the letter in the event the issuing bank fails to make proper payment. Having a letter confirmed is significantly different from having it "advised" by the U.S. exporter's bank. If the letter is advised, the U.S. bank is merely required to inform the exporter that the letter has been received by the U.S. bank—the U.S. bank does not add its own commitment to make payment under the letter. Consequently, having a letter of credit confirmed provides considerably more protection to a U.S. exporter than having it advised. Confirmation is commonly requested by U.S. exporters and is strongly suggested for most export transactions.

It is important to remember that the issuing bank is required to make payment upon receipt of the shipping documents; it has no obligation to verify the existence of the underlying goods themselves or the quality or quantity of such products. As a result, if the wrong goods or defective goods, even empty boxes, are shipped, the bank will pay as long as the documents are in proper order. Consequently, reliance on a letter of credit alone is insufficient for proper legal protection in a trade transaction. Parties should coordinate the use of a letter of credit with a properly prepared sales contract and related legal documents as well as make arrangements for inspection of the goods.

Remember, too, letters of credit can be revocable or irrevocable. A revocable letter of credit can be amended by the issuing bank at any time and consequently offers less protection than an irrevocable one. Exporters are strongly urged to insist on irrevocable letters of credit.

Most documentary letters of credit incorporate by reference a more elaborate set of rules and procedures that govern their use. This is the Uniform Customs and Practice for Documentary Credits, published by the International Chamber of Commerce. Users of letters of credit should be familiar with the provisions of the UCPDC and all the limitations, conditions, and other restrictions that apply when letters of credit are utilized.

Payment against collection documents. Under this method of payment, the seller ships his goods and simultaneously forwards his shipping documents through his bank to the buyer's bank. The shipping documents include the bill of lading for the goods, invoices, various consular and import documents, and the seller's draft drawn on the buyer. When the documents arrive at the foreign bank, the bank notifies the buyer. The buyer pays the amount owed to the seller, and the bank then releases the documents. When the buyer acquires the shipping documents, he can then take lawful title to the goods when they arrive; at the same time the bank forwards the funds to the seller. If the buyer does not make payment, the bank does not release the documents, and the buyer cannot obtain the goods.

This method has the advantage of assuring payment prior to the buyer obtaining the right to receive the goods. But it is less effective than a letter of credit since the seller must bear the cost of delivery of the goods to the original destination, and possibly the return voyage if the buyer does not make payment. In addition, as with letters of credit, the bank forwards documents only. There is no responsibility of the bank to assure that the proper quality and quantity of goods have been shipped.

Bank guarantees and stand-by credits. This method of payment is similar to the documentary letter of credit described above. As discussed, under documentary letters of credit, the bank's obligation to pay is triggered by the delivery of shipping documents, hence the term. But under bank guarantees and "stand-by" letters of credit, the bank's obligation to pay is triggered by the occurrence or failure of other events, such as the performance of a construction contract or repayment of an obligation. U.S. banks prefer to use documentary letters of credit in trade transactions rather than guarantees and similar instruments, but exporters may find foreign banks prefer to use guarantees, bonds, or stand-by payment instruments.

Consignment. As with domestic consignment, a U.S. seller will ship goods to its ultimate location, display them for sale, and receive payment when the goods are sold. This presents yet another method of payment and is a common way of conducting business in certain industries. As a practical matter, however, this method presents a significant payment risk for the exporter since he has no assurance of receiving ultimate payment. As a result, consignment should be avoided by exporters.

Open account. Once an exporter has established a positive working relationship with an overseas buyer, he may wish to begin dealing on an open account basis. Under this method, the exporter ships the goods, sends an invoice to the purchaser, and requests payment in 30 days, 60 days, 90 days, or what have you. This method poses considerable risk to the exporter, of course, since he ships his goods with no assurance of payment. As a practical matter, however, sellers frequently use this method in instances where the seller has an established and acceptable credit relationship with the foreign buyer or the seller is truly interested in concluding the sale. But, remember, unlike in domestic commerce, if the foreign buyer fails to pay the invoice, it is much more difficult to collect through international litigation than in a domestic lawsuit.

The Role of Documents in Trade Finance

The key intermediary in most trade payment transactions is the commercial bank. Banks cannot deal in merchandise—they generally deal in documents. Consequently, the reliability of the world trade payment system is based upon the use of standardized documents recognized and accepted at all financial institutions around the world. These include the bill of lading, commercial invoice, consular invoice, commercial draft, letters of credit, export and import licenses, insurance certificate, shippers export declaration, and certificate of origin. Familiarity with and proper use of these documents are essential for any person looking to structure a successful trade finance transaction.

Despite their variety and differences, a number of common principles can be applied to all of these documents. First, most of the documents have legal significance considerably greater than what is printed on their face. For example, a bill of lading frequently serves as both a document evidencing legal title to the ownership of the goods as well as a contract of carriage between the seller and an ocean or air carrier. These are special legal relationships based upon myriad statutes and case precedents not described in the text of the document. Similarly, most documentary letters of credit incorporate by reference the Uniform Customs and Practices for Documentary Credits, a complex set of rules regarding conditions for payment. Managers who utilize these documents in trade payment arrangements should be familiar with their full legal significance; improper use could subject them to considerable financial risk and legal liability.

Second, the legal significance of each document often changes from country to country based on the local law of the particular country. Consequently, parties should also be familiar with how such documents are customarily used in *international* commerce—reliance on U.S. legal definitions of such documents alone is insufficient.

Finally, improper use of some of these documents could lead to severe civil, even *criminal*, sanctions. For example, attempts to export products without export licenses issued by the U.S. Department of Commerce could lead to civil fines or imprisonment for a U.S. or foreign exporter. Similarly, attempts to export products into certain foreign nations without appropriate import licenses could subject the parties to possible fines or arrest.

Bottom Line

In conclusion, when structuring trade payment arrangements:

Check for accuracy. Because of the emphasis on documents rather than underlying goods in trade payment transactions, letters of credit and all related shipping documents should be properly and accurately prepared and carefully proofread; improperly prepared documents routinely cause delay in payments or more severe economic loss. Names and addresses of the parties, descriptions of the goods, amounts to be paid, and related items should all be confirmed and properly set forth.

Send a complete set of documents. It is essential that *all* documents required be forwarded to the bank for collection. If any document is missing from the list, the bank may not honor the letter of credit.

Assure irrevocability. Letters of credit and the like should be irrevocable, not subject to amendment by the issuing bank.

Assure confirmation. The exporter's U.S. bank should *confirm* rather than *advise* the foreign bank letter of credit; advising alone provides considerably less protection to the U.S. party.

Allow adequate time for approval. Be sure that all of the required documents are delivered as required, that adequate time is allowed for all necessary attestation and authentication procedures, and that your bank and other representatives have proper local relationships in the foreign country to expedite transmission and processing of the documents.

Know your buyer. Take adequate steps to learn about the party with whom you are dealing. Trade payment mechanisms generally only protect against risk of payment; they do not address quality, quantity, and other issues.

Use qualified experts. Use bankers, freight forwarders, customs brokers, and attorneys experienced with international business transactions and with the type and size of transaction you are conducting.

Coordinate with sales contract. Since the bank is under no obligation to ascertain conformance of the underlying goods, adequate provisions must be set forth in the sales agreement regarding quality, timing of delivery, return of defective goods, and so forth. In addition, adequate provisions must be made for inspection of the goods upon arrival. Trade payment mechanisms alone do not provide complete protection for most export transactions.

○ ○ ○

UPCOMING: Financing arrangements

THOMAS B. MCVEY is an international business attorney with the law firm McVey & Sherman in Washington, D.C.

WHARTON REPROGRAPHI

JOURNAL OF WORLD TRADE 1988, v. 22 no. 6.

Modern Methods of Processing Overseas Trade

Hugh M. Kindred*

Introduction

The carriage of goods by sea has commonly been conducted under ocean bills of lading, but in the last twenty-five years, or thereabouts, the goods in international trade have been carried in new containers bearing different labels. Commercially speaking, the containers are now twenty- and forty-foot metal boxes, while the documentary labelling has been redesigned and is about to be replaced. This article deals with the changes in the documentary processing of overseas trade and in particular reviews the legal consequences of displacing the bill of lading.

Once the universal document in foreign trade, the ocean bill of lading is today far from being the only choice of carriage agreement. Indeed, on some trade routes the bill of lading has already become virtually extinct. Its replacement is now official policy. Trade organisations such as the International Chamber of Commerce, the International Chamber of Shipping and the Economic Commission for Europe have gone on record with the recommendation that the use of bills of lading in ocean movements of goods should be avoided unless essential.[1]

The changes in the documentary practices of overseas trade raise a number of significant questions. Among those worth considering are the following: (1) Why is the billing of lading being replaced? (2) What are the new forms of shipping contracts? (3) What are the legal consequences of these changes? In this article each question is considered in turn, giving emphasis to the legal implications.

Replacement of the Bill of Lading

So great have been the changes in ocean transportation in the past twenty to thirty years that the shipping industry today operates in a completely different environment. Technological innovations have permitted enormous improvements in ship design, navigational equipment and operational efficiency. This has resulted in safer, speedier and more reliable shipping services. In addition to the improvements in ships, huge advances have been achieved in cargo handling through the introduction of unitization methods, especially containerization. The results are simplified and accelerated cargo movements. In fact, transit times on some containerized liner services have been cut so drastically that goods can now travel faster than their associated documents. For instance, on the North Atlantic route between Canada and Europe a container of goods is likely to reach the discharging port before its air-mailed bill of lading. Since the bill has to be surrendered to the carrier before delivery may be made, this documentary process is no longer convenient. Both the carrier and the cargo owner are put to extra cost and delay either to await the bill of lading or to secure release of the goods by other means.

---

* Faculty of Law, Dalhousie University, Halifax, Canada. This article is based on an address delivered at the annual meeting of the Canadian Transportation Research Forum, in Minaki, Ontario, May 24-28, 1988.

[1] e.g. E.C.E. Working Party on Facilitation of International Trade Procedures, *Measures to Facilitate Maritime Transport Document Procedures: Recommendation No.12.* UN/ECE/TRADE/WP.4/INF.61/1979.

Fortunately, complementary to the changes in shipping, methods of documentary processing have also been transformed. It is these new methods which have permitted the creation of novel forms of shipping contracts to replace the increasingly inconvenient bill of lading.

New Forms of Shipping Contracts

Speaking generally, the 1970s may be regarded as a period of simplification of trade documentation, and the 1980s as a time of introduction of electronic data processing.[2]

The number and variety of documents and forms associated with each sale and movement of goods became so great that it imposed an excessive cost on commerce. The first step was to reduce the number of forms. If a shipment of goods which might have required, for instance, thirty different forms could be moved and controlled with only eight revised documents,[3] there had to be a substantial saving in processing costs. The benefits of this order of simplification both in the numbers and the contents of trade documents is obvious.

Perhaps even more significant in impact has been the move towards documentary standardization made nationally and internationally. A seachange in documentary design has been achieved by the creation of look-alike forms for common trade and shipping functions that are suitable for use anywhere in the world. The common format also permits overlay, snapset and other one-run systems of cheaper document production. These economies have been achieved nationally through boards and councils for the simplification of trade procedures, such as SITPRO in the United Kingdom, SWEPRO in Sweden, SIMPRO in France and JASPRO in Japan. All these bodies are linked internationally through the Economic Commission for Europe Working Party on Facilitation of International Trade Procedures, which has promulgated such global standards as the U.N. Layout Key for Trade Documents.[4]

In application to shipping contracts, the first documentary improvements were directed to simplifying the bill of lading itself; in fact, short form and blank back bills have been in use for many years now. However the single most important documentary innovation for maritime transportation was, and is, the sea waybill.

---

[2] A brief history of the trade facilitation effort is provided by A. R. Elliott, *The Changing Face of Documentation in the Carriage of Goods* in D. J. Sharpe & W. W. Spicer *New Directions in Maritime Law 1984* Toronto: Carswell, 1985, at 193.

[3] Example provided by A. R. Elliott, *id.* at 197.

[4] For instance, this standard is the basis of the Canadian Trade Document Alignment System adopted by the Canadian General Standards Board in December 1982.

Sea waybills represent a relatively recent attempt to adopt the form of waybills well known in air transport and some land modes to ocean carriage. A unique feature of the ocean bill of lading is that, since it is a document of title to goods, it must be surrendered against delivery. While this function is useful when goods are to be traded while in transit, it is also the cause of much hindrance to commerce when the documents are late in arriving at destination. A sea waybill looks like a blank-back bill of lading and acts similarly as a receipt for the goods and as a contract for their carriage, but it cannot be negotiated as a document of title. Consequently, there is no need to present a sea waybill at destination in order to receive the goods. The consignee has only to identify himself to the satisfaction of the carrier as the party, or his agent, named in the waybill in order to have the cargo released to him. In this way the late arrival of documents does not delay delivery. Indeed, no document is required at destination at all. Hence, the use of sea waybills in place of bills of lading can expedite the delivery of goods by sea to the benefit of carrier and cargo owner alike.

These were the beneficial changes of the seventies, and even now in the eighties, they are still spreading in use and effect. But additionally this decade has witnessed the introduction of electronic data processing or EDP. This innovation will cause nothing short of a revolution in the documentary processing of overseas trade.

Applied to ocean transportation, the operational use of EDP is to be anticipated in the form of electrodocs. They are still so novel that the name "electrodoc" has not yet passed into general commercial usage,[5] but is a commendably suggestive nomenclature. Electronic communications have made possible the transmission of trade and transport data without documents. The message electronically transmitted is the electrodoc. Thus the same information that would be contained in a bill of lading may in principle be expressed in electronic impulses within a computer system and may be accessed by any authorised user with a connected and compatible terminal.

The widespread acquisition of computers for business office purposes makes electronic communications between traders, carriers, and their agents imminent. Already in overseas trade, trans-Atlantic liner companies have computer-linked their offices and all their port agents on both sides of the ocean. They no longer use documents intra-corporately, but transmit all their cargo manifests and instructions electronically. The next step is to begin concluding inter-corporate carriage agreements by teletransmissions, and this process awaits only the international standardization of existing technology. There have been a large number of pilot projects in this direction[6] and work is well advanced at the Economic Commission for Europe in drawing up standards of compatibility for trade data interchange.[7] The momentum for change is growing rapidly for it is evident to everyone involved that the instantaneous teletransmission of transport data will be infinitely more efficient, convenient and economic than paper documents such as bills of lading.

---

[5] The name is employed as a generic term in the draft C.M.I. *Uniform Rules for an Electronic Bill of Lading, Waybill or Other Transport Document* (1987).

[6] The North American grocery industry, for instance, is already operating with teletransmitted purchase orders and invoices. The universal product codes on food items in supermarkets are symbols of this development. The field of international trade is replete with acronyms of trial projects, such as CARDIS, Mercator, ODETTE, CEFIC and DISH.

Legal Consequences of the New Shipping Contracts

Since the new shipping contracts are intended as replacements for the bill of lading, it is essential that they fulfil the functions expected of them by all the participants in overseas trade. Part of those expectations is an assumption that the new ways of handling trade data will be recognized and enforced at law. Unfortunately, when disputes break out the present law may not be so conducive to their resolution as the commercial parties would like. The new shipping contracts have two functions. They are intended to act as receipts for the goods shipped, and also to provide evidence of the agreement for carriage. In each case legal doubts may be raised about the accomplishment of these functions.

A. As Receipts

The carrier's receipt for the goods being shipped has traditionally been supplied by the data on the face of the transport document. The purpose of a receipt is to provide a record of the shipper's goods in possession of the carrier. It details the quantity, marks, description, weight or volume, and the apparent condition of the cargo, as well as the identities of the shipper, the carrier and the receiver. It may be used as evidence by the cargo owner to prove his right to the goods. As short form agreements and sea waybills provide a documentary face similar to the standard ocean bill of lading, they are able to function as receipts in precisely parallel ways without further consideration. Electrodocs cannot. Since the medium of their communication is not paper but electronic impulses, they do not display carriage data in a documentary form.

Computers, however, are particularly good at creating, keeping and transferring records amongst different data bases. There are no technical obstacles to teletransmitting normal transport data so there should be no difficulty in employing electrodocs to respond to the shipping industry's need for a receipt. The main concern, therefore, is whether the law will recognise and accept an electrodoc as an effective "electronic receipt".

Though the law is concerned with acts and events, very often it operates upon them by way of their form. Thus in trade matters many of the legal controls on commercial dealings are expressed in terms of the documents in which they have traditionally been performed, but any law written according to the paper medium will obviously obstruct the use of electronic transmissions. Unfortunately paper-tied laws are plentiful at this time.

---

[7] See E.C.E. Working Party on Facilitation of International Trade Procedures, *Electronic Data Interchange for Administration, Commerce and Transport (EDIFACT)*, ISO Standard 9735, UN/ECE/TRADE/WP.4/R.513 (Dec. 1987).

Canadian law in this regard is typical of the experience in common law countries generally. These nations have enacted similar legislation and followed each other's judicial precedents in the interest of maintaining a measure of uniformity that is natural, if not essential, to the orderly conduct of overseas trade. Examples of the different ways that Canadian trade laws are linked to the paper medium are not hard to find. Some refer to a specific document; thus the Bills of Lading Acts[8] name the bill of lading and thereby exclude other forms of carriage contract. Other laws call for a written form or notice; for instance the carrier must provide a notice in writing under the Canada Shipping Act in order to retain his lien on landed goods.[9]

Sometimes the law specifies requirements in such a way that it imposes the use of a paper document; thus the Hague Rules, the international uniform rules for ocean transportation, grant the shipper of goods a right to demand from the carrier a statement of the particulars of the cargo in a bill of lading.[10] At other times the law implies the existence of a paper document; for instance the Customs Act demands the production of an "original" record.[11] The statute books are evidently full of restrictive laws that will impede the use of electrodocs. In the end, reform of all commercial laws referring expressly or implicitly to paper transactions will be necessary.

B. As Contracts of Carriage

In contrast to the receipt function of the face of the transport document, the contract of carriage is usually evidenced by the fine print on the back. Few commercial people spend time reading the lengthy and plentiful conditions of carriage, leaving that to their lawyers, so it has to be observed that, despite their apparent comprehensiveness, they may be open to a number of legal doubts. Among those in need of examination are the following questions: (1) What are the terms of the contract? (2) Who are the parties to the carriage? (3) How is the shipping agreement proven?

1. TERMS OF THE CONTRACT

Neither sea waybills nor electrodocs express the terms of the carriage contract, rather they attempt to establish an agreement by reference to the carrier's standard trading conditions found elsewhere. The conditions are invoked by means of an incorporation clause placed amongst the carriage data. The critical question is what legal force such incorporation clauses may have. There seem to be two problems: one concerns the reference to the carrier's standard terms and the other involves the application of international laws such as the Hague and Hague/Visby Rules.

---

[8] e.g. R.S.C. 1970, c. B-6.

[9] R.S.C. 1970, c. S-9, s. 667.

[10] See Carriage of Goods by Water Act, R.S.C. 1970, c. C-15, Schedule, art. III.

[11] S.C. 1986, c.1., s. 40 and SOR/86-1011.

(a) *Carrier's Standard Terms*

The crux of this problem is whether the incorporation of the carrier's standard conditions by reference is sufficient to bind other parties. It might be doubted whether legal consequences could attach to a merchant who was ignorant of the contractual terms allocating the risks associated with the movement of his goods. Although the law looks for an agreement to found a contract, it is prepared to read a concurrence of intentions out of a situation. Thus, if the shipper and carrier have had previous dealings, carriage according to the carrier's standard conditions may be inferred. If the parties are involved in their first transaction together, but the cargo owner is by profession a shipper, exporter or importer, then the standard conditions of the trade may be implied.[12] Hence the lack of notice of standard terms and conditions is usually irrelevant between contracting parties who are participants in the shipping business.

An exception is made when the incorporated terms are unusual in the trade. Then notice to the merchant is essential and the kind of notice required is relative to the degree of surprise and onerousness of the term being relied on.[13] The requirement is founded in reasonable business efficacy. "It is impractical for a shipper to be compelled to make a detailed study of all the fine print clauses of the carrier's regular bill on each occasion before it ships out a package".[14]

(b) *The Hague Rules*

The Hague Rules, or their amended version the Hague/Visby Rules, are internationally agreed rules for the carriage of goods by sea. They create minimum standards of responsibility for ocean cargoes and thus provide a settled, uniform distribution of risks and liabilities in overseas trade. The Hague Rules have the force of law within the scope of application established by the implementing legislation in each country. In Canada,[15] the Carriage of Goods by Water Act[16] mandates the Hague Rules' operation in contracts of carriage covered by bills of lading for goods being exported by sea. The statutory effect of the Rules is to override any express term of the carriage contract that diminishes the carrier's liability established by them.[17]

---

[12] *George Mitchell Ltd. v. Finney Lock Seeds Ltd.*, [1983] 1 Lloyd's R. 168 (C.A.), aff'd. [1983] 2 All E.R. 737 (H.L.); and *British Crane Hire v. Ipswich Plant Hire*, [1974] 1 All E.R. 1059 (C.A.). See also E. MacDonald, *Incorporation of Contract Terms by a 'Consistent Course of Dealing'* (1988), 8 Legal Studies 48.

[13] *Crooks v. Allan* (1879), 5 Q.B.D. 38; and *Spurling v. Bradshaw*, [1956] 1 W.L.R. 461.

[14] Per Anderson J. in *Encyclopaedia Britannica Inc. v. The Hong Kong Producer and Universal Marine Corp.*, [1969] A.M.C. 1741, at 1750.

[15] As in the United States, and similarly for the Hague/Visby Rules in the United Kingdom.

[16] R.S.C. 1970, c. C-15, s.2.

[17] *Ibid.*, Schedule art. III (8).

Even when the Hague Rules do not apply by statute, carriers frequently incorporate them by reference, often as a clause in their standard conditions. For instance, the application of the Rules might be extended to containers carried on deck, or to goods in the carrier's charge before loading or after discharge. The carrier benefits by his exposure to the risks of loss being limited to a low level of liability per package.[18] When contractually incorporated, the Hague Rules have no statutory priority but, in principle, are coequal with the carrier's other conditions. Then conflicts between the Rules and the carrier's terms are possible and problems of contractual interpretation may arise.

Courts in different coutries seem to have adopted different approaches to these problems and to have reached varied results. In Canada it has been said "the maker of the bill ought not to benefit from any ambiguity of his own creation".[19] Thus express clauses of the carrier's making that are intended to lessen his liability under the Rules should be construed against him. For instance, in the case of *The Doroty*[20] the carrier's clause that defined the rate for assessing liability according to customary freight units (an American practice) in place of shipping units (the Canadian interpretation of the Hague Rules) was disallowed. The carrier had to pay at the Hague Rules rate of liability for each of the rotting fish he delivered.

In the United States, one line of cases, at least, follows the principle of contract construction that a more specific provision should be advanced over a more general term. In shipping situations, this approach will generally result in a preference for the carrier's express conditions over the Hague Rules since they are likely to have been drafted very precisely to take account of the existing Rules.[21] The English courts' approach, arguably, is to resist the incorporation of those articles of the Hague/Visby Rules which are inconsistent with the express terms in the carriage document.[22] The carrier's conditions are preferred, perhaps, beause their deliberate expression better shows the parties', or at least the carrier's, contractual intentions.

These approaches expose a sharp distinction in judicial policy. The Canadian readiness to sever the carrier's terms that are inconsistent with the Hague Rules, which he has also chosen to invoke, respects the public policy supporting those Rules. They are meant to be internationally determined, uniform, minimum standards of conduct and responsibility between carriers and cargo owners. By promoting the parties' inferred intentions, the American and English perspectives uphold the private law principles of freedom and integrity of contract. The weakness of this stance may be that the freedom of contract being supported is really the power of the carrier to dictate unilaterally the terms of carriage. The imbalance in the contracts resulting from similar lopsided bargaining power was precisely the reason why the Hague Rules were originally concluded more than sixty years ago.

---

[18] *Ibid.*, Schedule art. IV (5).

[19] Per Dubé in *Atlantic Consolidated Shipping Co. Ltd. v. The Doroty*, [1979] 1 F.C. 283, at 297, aff'd [1981] 1 F.C. 783.

[20] *Ibid.*

[21] e.g. *Pannell v. U.S. Lines*, [1959] A.M.C. 935, 263 F. 2d 497; and *Commonwealth Petrochemicals Inc. v. Puerto Rico*, [1979] A.M.C. 2772.

[22] See *Anglo-Saxon Petroleum Co. v. Adamastos Shipping Co.*, [1958] 1 Lloyd's R. 73 (H.L.).

The use of sea waybills poses a further problem respecting the operation of the Hague Rules. Since the Rules themselves are expressed to apply to contracts for the carriage of goods by sea "covered by a bill of lading or any similar document of title",[23] there is considerable doubt whether they ever apply to waybills with statutory force and priority. Technical legal arguments can and have been made that sea waybills might be included as documents of title, or that other substantive Rules can be interpreted so as to circumvent the apparent limitations of this language,[24] but none are guaranteed of success.

These arguments have not been tested in the courts in part because sea waybills commonly incorporate some version of the Hague Rules along with the carrier's standard conditions. Consequently, the shipping community tends to think there is no problem, but, as previously shown, there is considerable uncertainty about how the Hague Rules will operate contractually. Further, carriers are not bound to incorporate the Hague Rules: on the contrary, they may choose to employ a jurisdiction clause that selects as the governing law and courts the legal system of some country where the Hague Rules do not operate.[25]

The uncertainty over the statutory impact of the Hague Rules also exists with regard to electrodocs. The making of the problem seems to have been inadvertent. The new shipping contracts were not in use when the Hague Rules were written and so the Rules were drafted in terms of bills of lading. However, there is no reason why the new forms should not be subject to the Rules since they are intended to play the same role with respect to the carriage contract as the bill of lading.[26] This objective could be achieved by an amendment to national implementing legislation, such as the Carriage of Goods by Water Act in Canada, so as to extend the Hague Rules to all kinds of shipping contracts, documentary and electronic,[27] but so far no such proposal has been presented to any Parliament.[28] In the meantime, the distinct commercial advantages of the new shipping contracts to both carriers and merchants will require the courts to exercise considerable creative ingenuity in order to continue applying the same balance as before in the allocation of marine risks and cargo liabilities when disputes arise.

---

[23] *Supra* n. 17, Schedule art. I(b).

[24] See Comments of the Committee on Carriage of Goods of the Maritime Law Association of the United States regarding C.M.I. *Uniform Rules for Incorporation into Sea Waybills or Other Similar Documents* at 2-4 (January. 1988); and W. Tetley, *Waybills, the Modern Contract of Carriage of Goods by Sea* (1983), 14 *Jo. Mar. L. & Comm.* 465, at 471–473.

[25] See e.g. *Argo Company of Canada Ltd. et al. v. Argonaut Marine Inc. (The Regal Scout)* (1983), 148 D.L.R. (3d) 412 (F.C.T.D.). But courts do not always enforce such choice of law clauses.

[26] Cp. the Hamburg Rules which are expressed to apply to contracts of carriage by sea without regard to form. See arts. 2, 1(6) and 14(3).

[27] Always excepting charter parties.

[28] The English Parliament, in the course of adopting the Hague/Visby Rules, went some way when it enacted the curious provision that they shall have the force of law in a sea waybill if it "expressly provides that the Rules are to govern the contract as if the receipt [waybill] were a bill of lading". See U.K. Carriage of Goods by Sea Act 1971, 1971, c. 19, s. 1(6).

2. PARTIES TO THE CARRIAGE

The shipping contract is agreed between the shipper and the carrier for the purpose of carrying the cargo to the consignee. There is no doubt that the shipper and the carrier are parties to the contract since they made it, but the position of the consignee, as the ultimate beneficiary of the carriage, is less certain. The usual commercial intent is that the consignee at some point in time during the transit, frequently when he pays for the goods, shall take over the contract with the carrier in place of the shipper. This substitution of the consignee for the shipper as the contracting party on behalf of the cargo has important effects. It grants the consignee rights of control over the goods, either to demand their delivery or to redirect their movement by the carrier. It also binds the consignee directly to the carrier for any outstanding freight or other charges payable. The ability to transfer the shipper's interests in the carriage contract is therefore crucial to empower the consignee with respect to the goods.

As it happens, the common law courts long ago decided that the law did not recognise attempts to transfer contractual rights to strangers to the original agreement,[29] apparently because they were regarded as too personal in nature to suffer the change. Thus, the shipper could not effectively transfer the carriage rights, even though he had separately sold the goods themselves, to the consignee. Such a consignee would be in the untenable position of being the owner of a cargo over which he had no lawful control.

Fortunately the legislatures came to the rescue. As long ago as 1855 the British Parliament passed the original Bills of Lading Act[30] which has since been copied in other jurisdictions, including Canada.[31] The effect of these acts is to permit the assignment of the rights and responsibilities of a shipper arising under a bill of lading. The consignee under a bill of lading is thus protected by the assurance of becoming a full party to the carriage contract. But the Bills of Lading Acts, as their name implies, operate only on bills of lading and do not extend to other types of shipping contracts. So consignees under sea waybills and electrodocs will be exposed to the risk that they might not be able to direct delivery of their goods or to sue the carrier for misdelivery. Conversely, the carrier might discover that he cannot pursue the consignee upon the terms of the contract for payment of freight or charges, and may not rely on the limitation of liability, including the Hague Rules, that he deliberately incorporated into the shipping agreement.

In the light of these legal risks, some carriers have added a term in their sea waybills by which the parties purport to grant themselves the power to transfer their interests contractually.[32] These clauses rely on the basic principles of contract law and equity that the intentions of the parties should be given effect, but there is no certainty that the legal proscription of assignment can be so easily overridden. An alternative legal argument, which has met with some success,[33] is to imply a separate contract between the consignee and the carrier, on exactly the same terms as the original shipping agreement, as a result of the acts of delivery of the goods and payment of the freight charges.

---

[29] *Thompson v. Dominy* (1845), 14 M. & W. 403, 153 E.R. 532.

[30] 18 and 19 Vict., c. 111.

[31] *Supra*, n. 9.

[32] e.g. the standard conditions for sea waybills drafted by the General Council on British Shipping.

The legal uncertainty surrounding such a basic commercial function as the transferability of the carriage contract remains unsatisfactory and demands legislative reform. So much so, the *Comité International Maritime (C.M.I.)*, the original author of the Hague Rules, has struck an international subcommittee for this very purpose. Draft Uniform Rules for Incorporation into Sea Waybills or Other Similar Documents and draft Uniform Rules for an Electronic Bill of Lading, Waybill or Other Transport Document are now under active discussion. They are intended to resolve the problems raised here, amongst others, by the provision of model rules which will be presented for approval by the C.M.I. at its assembly in 1990, and thereafter will be available internationally for adoption and enactment by national legislatures.

3. PROOF OF THE CARRIAGE CONTRACT

The shipping agreement itself is generally regarded as the best evidence of the terms of carriage.[34] Under the Hague Rules, a bill of lading is *prima facie* proof of receipt by the carrier of the goods it describes until rebutted by the opposing party.[35] The new shipping contracts in documentary form may safely constitute proof on the same basis, but uncertainties surround electrodocs because they lack documentary character. At bottom, the question is how the law will regard electronically-generated evidence. Two aspects of the electrodoc seem particularly troubling: the first involves the authentication of the carriage contract, and the second concerns the proof of its contents.

(a) *Authentication*

The usual form of authentication of a documentary transaction such as an agreement of sale or a contract of carriage is by signature. Typically the signature is an autograph, but it may be impressed with a rubber stamp or a corporate seal, or by perforating or embossing the document. The purpose of the signature is to acknowledge authorship of and responsibility for the contract.[36] It is executed by deliberately making a personal mark.[37] In other words, the signature personalizes the document. Computer systems permit the application of comparable individualization to an electrodoc,[38] and permit the use of unique personal codes with at least as much security as the risk of forged signatures.[39] The commercial need for authenticity of an electrodoc being thus satisfied, the question outstanding is whether the law will accept such "electronic signatures".

---

[33] See *Brandt v. Liverpool, Brazil and River Plate Steam Navigation Co. Ltd.*, [1924] 1 K.B. 575; and *The Aramis*, [1987] 2 Lloyd's R. 58 (Q.B.).

[34] *Sewell v. Burdick* (1884), 10 App. Cas. 74, at 105; *Crooks v. Allan* (1879) 5 Q.B.D. 38; *The Ardennes*, [1951] 1 K.B. 55.

[35] *Supra* n. 17, Schedule art. III(4).

[36] *Nasmith Co. v. Alexander Brown Milling Co.* (1905), 9 O.L.R. 21.

The courts have not taken exception to any form of authentication of documents, even when the imprint is indecipherable.[40] The critical feature is that the act of marking the document is done with intent to authenticate it.[41] In other words, any mark is enough as long as it is made with intent to sign. By analogy, an electronic signature ought to be equally acceptable.[42] An electrodoc may be implanted with a unique electronic code which, in a sense, is no more than a particularly sophisticated mechanical mark. Since the code is transmitted in the electrodoc for the purpose of individualising the enclosed message, there is hope the courts will consider it a sufficient authentication, provided it can be proved.[43]

(b) *Proof of Contents*

Proof of a shipping contract expressed in an electrodoc cannot be made simply by production of the item on which it is inscribed, as in the case of documentary agreements. While the contents of a bill of lading or a sea waybill may be read directly from the document, computer tapes and diskettes provide no intelligible evidence. Electronically stored records are machine readable only and print-outs are at best copies, not truly original evidence. Worse, the output may be the result of processing the stored data into quite different material and not the reproduction of copies at all.[44] The law, consequently, faces a novel problem concerning the value to be ascribed to an electrodoc evidence to prove the contents of the shipping contract.

At the outset, a basic legal distinction must be recognised between the admissibility and reliability of evidence. In general, only direct and relevant evidence is admissible. Even when admissible, evidence must be weighed for its credibility,. The

---

[37] *Caton v. Caton* (1867), 11 L.R. 2 H.L. 127.

[38] See the proposal for the use of alphanumeric codes in the draft C.M.I. *Uniform Rules for an Electronic Bill of Lading, Waybill or Other Transport Document* (1987).

[39] See E.C.E. Working Party on Facilitation of International Trade Procedures, *Authentication of Trade Documents: Recommendation No. 14*, para. 23 (1979).

[40] *Cf. re United Canso Oil & Gas Ltd.* (1980), 41 N.S.R. (2d) 285 (N.S.S.C.).

[41] *Caton v. Caton, supra* n. 38.

[42] *Cf.* the Hamburg Rules art. 14(3) which would permit the signature on a bill of lading to be made electronically if not inconsistent with the law of the country where it is issued.

[43] The electrodoc may run up against a legal requirement that its authentication must be in written form. This obstacle is comparable to problems over the form of electrodocs as receipts, discussed earlier, and will require similar legal reforms. The Customs Cooperation Council has been much concerned with this possibility and has called for international action to give electronic signatures the same status as holographic ones.

[44] See C. Tapper, *Evidence from Computers (1974)*, Ga. L.R. 562.

[illegible] of computer operations the [illegible] and [illegible] been [illegible] and [illegible] but the [illegible] the reliability of [illegible] the internal [illegible] organizations.

[illegible]

[illegible]

[illegible]

9

WHARTON REPROGRAPHICS

SCHOOLS BRIEF

Delivering the goods

The vast expansion in international trade owes much to a revolution in the business of moving freight. The fifth in our series of briefs on globalisation explains how it happened

IF SHOP windows everywhere seem to be filled with imports, there is a reason. International trade has been growing at a startling pace. While the global economy has been expanding at a bit over 3% a year, the volume of trade has been rising at a compound annual rate of about twice that. In 1996 some $5.2 trillion of goods was sent from one country to another, up from $2 trillion a decade earlier. Foreign products, from meat to machinery, play a more important role in almost every economy in the world, and foreign markets now tempt businesses that never much worried about sales beyond their nation's borders.

What lies behind this explosion in international commerce? The general worldwide decline in trade barriers, such as tariffs and import quotas, is surely one explanation. The economic opening of countries that have traditionally been minor players in the world economy, such as China and Mexico, is another. But one force behind the import-export boom has passed all but unnoticed: the rapidly falling cost of getting goods to market.

In the world of trade theory, shipping costs do not matter. Goods, once they have been made, are assumed to move instantly and costlessly from place to place. The real world, however, is full of frictions. Exporting Italian steel to Britain may be good business if freight costs £20 ($34) a tonne and impractical if it costs £100 a tonne. Cheap labour may make Chinese clothing competitive in America, but if delays in shipment tie up working capital and cause winter coats to arrive in April, trade may lose its advantages. Freight costs, in short, can have a huge impact on both the overall volume of trade and on individual countries' trade patterns.

Neighbourly 1

Leading world exporters, 1996

	Biggest export market*
United States	Canada (21.3)
Germany	EU (56.4)
Japan	US (27.5)
France	EU (62.6)
Britain	EU (52.7)
Italy	EU (55.4)
Netherlands	EU (78.1)
Canada	US (82.3)
Belgium-Lux.	EU (70.4)
China	Hong Kong (21.8)
South Korea	US (16.7)
Singapore	US (18.4)
Taiwan	Hong Kong (39.6)
Spain	EU (79.0)

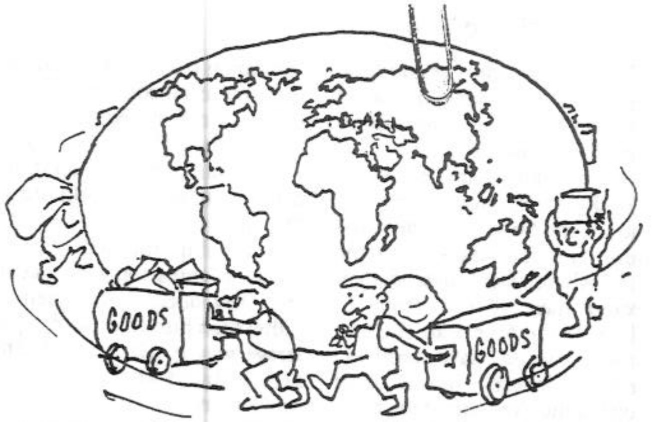
Source: IMF *% of total exports

They are a far less daunting obstacle than they used to be. This reflects two notable economic trends. First, the world economy has become far less transport-intensive than it once was. Second, the transportation industry has changed in remarkable ways, making it far cheaper and easier to ship goods around the world.

Brawn and brains

At the turn of the 20th century, agriculture and manufacturing were the two most important sectors almost everywhere. Together, they accounted for about 70% of total output in Germany, Italy and France, and 40-50% in America, Britain and Japan.

International commerce was therefore dominated by raw materials, such as wheat, wood and iron ore, or processed commodities, such as meat and steel. In 1900 "crude materials" and "crude food" made up 41% of America's exports, by value, and 45% of its imports. The story in Britain was similar. Basic materials and foodstuffs, both processed and unprocessed, accounted for 75% of imports and 20% of exports. These sorts of products are heavy and bulky. The cost of transporting them is relatively high, compared with the value of the goods themselves, so transport costs had much to do with the volume of international trade.



Countries still trade disproportionately with their geographic neighbours (see chart 1). Over time, however, world output has shifted into goods whose value is unrelated to their size and weight. Finished manufactured products, not raw commodities, dominate the flow of trade (chart 2). And thanks to technological advances, such as light-weight composites to replace steel and microprocessors to do the job of huge control panels, manufactured goods themselves have tended to become lighter and less bulky. As a result, less transportation is required for every dollar's worth of imports or exports.

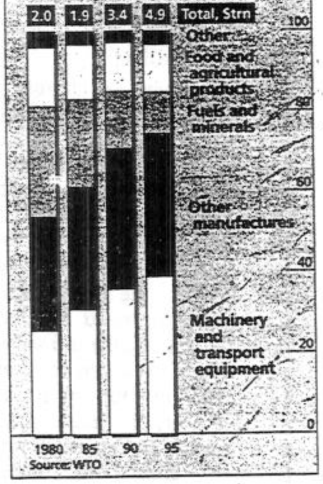
To see how this influences trade, consider the business of making disk drives for computers. Most of the world's disk drive manufacturing is concentrated in South-East Asia. This is possible only because disk drives, while valuable, are small, light and so cost little to ship. Computer manufacturers in Japan or Texas will not face hugely bigger freight bills if they import drives from Singapore rather than buying them domestically. Distance therefore poses no obstacle to the globalisation of the disk-drive industry.

This is even more true of the fast-growing information industries. Films and compact discs cost little to transport, even by aeroplane. Computer software can be "exported" without ever loading it on to a ship or a cargo plane, simply by transmitting it over telephone lines from one country to another. In such cases, freight rates and cargo-handling schedules become insignificant factors in deciding where to make the product.

The trend towards less transport intensity is clearly visible in trade statistics. The average dollar value of a unit of exported manufactured goods rose at an annual rate of 2% from 1990 to 1995, according to the World Trade Organisation. As that unit of exports is worth more and more, the cost of shipping it matters less and less. Businesses can locate based on other considerations, such as the availability of labour and good airline service for peripatetic employees, while worrying less about the cost of delivering their output.

Value added 2

World exports, % of total



Source: WTO

At the same time, the cost of shipping itself has fallen sharply. In many countries, deregulation has helped to drive the process along. But behind the scenes, a series of technological innovations, known broadly as "containerisation" and "intermodal transportation", has led to swift productivity improvements in cargo-handling—and, in the process, has lowered one of the biggest obstacles to trade.

Forty years ago, the very physical process of exporting or importing was arduous. Imagine a European textile manufacturer wanting to sell its wares in America. First, at the company's loading dock, workers would have lifted bolts of fabric into the back of a lorry. The lorry would have headed to a port and unloaded its cargo, bolt by bolt, into a dockside warehouse.

As a ship prepared to set sail, longshoremen would have removed the bolts from the warehouse and hoisted them into the hold, where other longshoremen would have stowed them in place. At the American end, the process would have been reversed: dockers would have hoisted the cloth out of the hold, carried it to storage and then, eventually, placed it into a lorry that would take it to its destination.

Simply arranging this sort of shipment was a complex task. The repeated cargo-handling and storage would have been expensive, and the cloth would have spent days, or even weeks, sitting in warehouses. Likely as not, portions of the shipment would have been damaged or stolen along the way. Add in high customs duties and importing was a costly affair.

Smoother sailing

That first began to change in 1955, when an American road-hauling magnate named Malcom McLean hit upon a more efficient way to ship goods. Under his original scheme, a lorry body, wheels and all, was unhitched from the driver's cab and lifted onto the deck of a ship, eliminating the need for dockworkers to handle the individual items inside the cargo compartment. This method soon gave way to the use of metal containers that could be separated from the lorry's chassis and wheels. With the chassis left at dockside, the containers could be stacked several high aboard the ship.

Mr McLean's first ship, an oil tanker called the *Ideal-X*, made its initial voyage from New York to Houston in 1956 with lorry bodies on its deck. In 1965 ships with containers on board began crossing the Atlantic. Ports were forced to adapt, as narrow piers in urban centres were replaced by large docks on the edge of town, alongside patios to hold thousands of containers.

The container crane, a new invention, made it possible to load and unload containers without capsizing the ship. The adoption of standard container sizes allowed almost any box to be transported on any ship. By 1967 dual-purpose ships, carrying loose cargo in the hold and containers on the deck, were giving way to all-container vessels that moved hundreds—and now thousands—of boxes at a time.

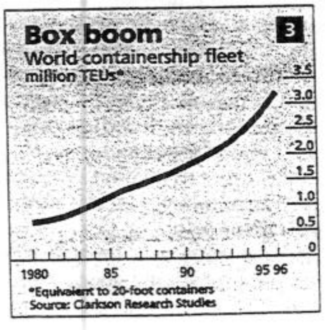
Containerisation had huge consequences for world trade. The most obvious was that the cost of shipping fell precipitously, as ships could be loaded by a few dozen longshoremen rather than hundreds, and as pilferage was much reduced. The need to build wooden crates to protect individual items was eliminated, making it feasible to ship consumer goods such as toys and stereo systems halfway around the world. International shipping capacity soared (chart 3), driven by large increases in the volume of goods shipped.

Mixing modes

The shipping container transformed ocean shipping into a highly efficient, intensely competitive business. But getting the cargo to and from the dock was a different story. National governments, by and large, kept a much firmer hand on lorry and railroad tariffs than on charges for ocean freight. Almost everywhere, rates were either set by state bureaucrats or were subject to their approval. New companies could enter the freight business only with great difficulty, and were subject to tight restrictions.

This started changing in the mid-1970s, when America began to deregulate its transportation industry. First airlines, then road hauliers and railways, were freed from restrictions on what they could carry, where they could haul it and what price they could charge. Lorries were no longer forced to run empty because they were licensed to carry goods on only one leg of a round-trip journey. Railways were no longer obliged to maintain unprofitable branch lines, but could focus on moving freight in large volumes over long distances.

Big productivity gains resulted (chart 4). Between 1985 and 1996, for example, America's freight railways dramatically reduced their employment, their trackage, and their fleets of locomotives and freight wagons—while increasing the amount of cargo they hauled. Europe's railways have also shown marked, albeit smaller, productivity improvements. Road hauliers have also been able to improve utilisation of their fleets, thanks to measures ranging from satellite tracking of vehicles, which enables more precisely timed pick-ups and deliveries, to soliciting back-haul cargoes to fill vehicles that formerly returned empty to their home bases.

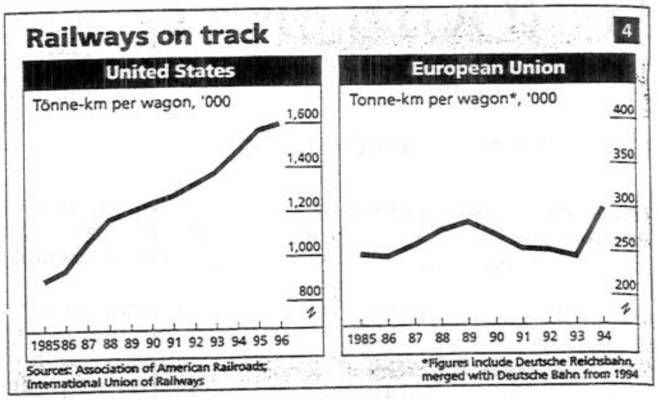

Box boom 3

World containership fleet

million TEUs*

*Equivalent to 20-foot containers

Source: Clarkson Research Studies


Railways on track 4

Sources: Association of American Railroads; International Union of Railways

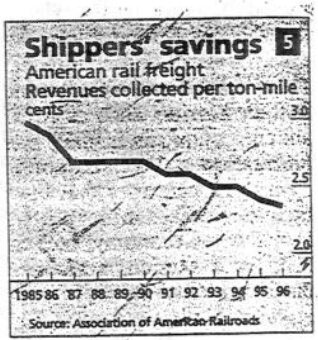
*Figures include Deutsche Reichsbahn, merged with Deutsche Bahn from 1994

These advances have benefited shippers, as tariffs for both road freight and rail freight have fallen sharply (chart 5). This made it feasible for companies located well inland to rely on imported components and to export, giving world trade yet another push. Impressed by the results, Japan and many emerging economies began to deregulate freight transport in the 1990s.

This freight revolution accelerated in the 1980s, as deregulation and new technology broke down the boundaries between different modes of transportation. For the first time, a manufacturer in, say, South Korea could ask a shipping line to deliver its exports to the American mid-west. The shipping line might strike a deal with a railway to haul the container from Los Angeles to Chicago; hire a road haulier to transport it from Chicago to South Bend; take responsibility for meeting delivery schedules at every stage of the journey; and send a single invoice for the entire shipment.

Intermodalism has given rise to cargo companies, such as FedEx and UPS, which specialise in using a combination of aircraft and lorries to deliver freight fast. It has led railway companies to build tracks at dockside, so containers can be moved directly from ships on to trains. And it has led to partnerships that would have been unthinkable a decade ago.

In Chicago, for example, lorries can pull into a railway marshalling yard, where overhead cranes can lift their containers on to waiting flat-car trains at the rate of one every 30 seconds. Two days later, having stopped en route only to change crews, the train arrives at a precise time in Los Angeles, nearly 2,000 miles away. Such careful management is squeezing yet more costs out of the transport system, making trade all the easier.


Shippers' savings 5

American rail freight

Revenues collected per ton-mile

cents

Source: Association of American Railroads

In America, the period of huge productivity gains in transportation may be almost over after two decades of deregulation. But in most other countries, the process still has far to go. Although trade itself has been liberalised, transportation has been slow to follow. State ownership of railways and airlines, regulation of freight rates and toleration of anti-competitive practices, such as ocean-shipping cartels and cargo-handling monopolies, all keep the cost of shipping unnecessarily high and deter international trade. Bringing these barriers down would help the world's economies grow even closer.

10

WHARTON REPROGRAPHICS

Geographical shifts in the competitive strength of mineral production since 1960, and their causes

Phillip Crowson

The strong influence of resource endowments, the depleting nature of ore extraction, and relative costs on patterns of location of mineral production are illustrated by reference to copper. Determined managerial action can offset seemingly immutable trends in competitiveness as the experience of the US copper industry in the 1980s demonstrates. Changes in the nature and pattern of demand are perhaps less obvious but important factors, again illustrated by copper. Evolving shares of world production of six other minerals are described to demonstrate other determinants of location. These include particularly transport and energy costs, and a wide range of government policies.

The RTZ Corporation PLC, 6 St James's Square, London SW1Y 4LD, UK.

This paper analyses geographical shifts in the location of mineral production and in competitiveness over the past 30 years.

The experiences of this period really need to be seen in a much longer time perspective. They are no exception to the underlying geological imperative that finite ore deposits will eventually be depleted by continuous mining. The geographical centre of gravity of mineral production has shifted throughout recorded history as resources have become exhausted, and as industry's needs have changed. Patterns of location have largely been the result of creative tension between various demand based factors, and the effects of supply developments. This interplay has been influenced by changes in transport systems and costs, and further modified by political, social, institutional and economic conditions. An analysis of locational change in the mining industry which ignores any of these aspects is liable to be lopsided and incomplete.

The techniques and cost structures of minerals extraction are very similar no matter what metals or minerals are being mined. The relative economics of mining *per se* in one region rather than another will be largely governed by national policies and economic conditions. These are superimposed on the underlying geological and geographical factors which determine the location of exploitable ore deposits. A further dimension comes from the structure of the market for each individual mineral product. One way of looking at these different dimensions is to treat them as the various single coloured images which combine to make a colour photograph. All contribute to relative competitiveness in any particular region or product, and the picture is incomplete without any one.

The initial sections of this paper use some examples from the copper industry to illustrate several general features of location in the mining industry. These include the crucial importance of mineral endowments and the depleting nature of ore extraction in shaping the industry's

0301-4207/92/040252-15 © 1992 Butterworth-Heinemann Ltd

geography. Levels of operating costs and their changing relationships, determined both by inherent features and by policy, are important driving forces. So too, and perhaps less obviously, are the location and nature of markets. Trends in competitiveness are not immutable, and they can be reversed by determined managerial actions, as witnessed by the US copper industry in the second half of the 1980s.

The messages about competitiveness drawn from the examination of copper are applicable to all minerals, whereas some other influences are not as widely applicable. The latter sections of the paper briefly discuss changing relative shares of world production of several minerals over the past 30 years to illustrate some of these influences. The discussion is only of the mining stage, rather than of downstream processing, however important that may be. Thus, the location of bauxite mining is examined, but not that of aluminium smelting. Six minerals, bauxite, gold, iron ore, nickel, phosphate rock and zinc, are discussed in alphabetical order to draw out the main features of their changing patterns of location. While some of the influences overlap, each product has been chosen to highlight different aspects of location. Finally, a concluding section draws together the main threads.

The lack of footnotes and the inclusion of only a few references may appear strange to academic readers. They are all too often a substitute for reasoned analysis and designed mainly to demonstrate the author's scholarship and depth of reading. In industry the analysis and conclusions stand or fall on their own merits. The views and opinions expressed in this paper are entirely personal, being drawn from the author's extended experience in the international minerals industry. Much of the analysis is based on the RTZ Corporation's databases on all aspects of the industry.

Ore depletion: the example of copper

That economic deposits of minerals are unevenly distributed throughout the globe is a statement of the obvious. Perhaps equally important is the depleting nature of minerals extraction. It is this feature, and its converse, that mines tend to be worked as long as they contain further economically extractable ore, which distinguish mining from other industries. At any one time the mining industry consists of successive generations of mines. Each generation will initially embody the techniques, and its location reflects the relative competitiveness of its time. Figure 1 peels back successive onion skins to show the contributions of several generations of mines to the Western world's copper production over the past 20 years.

The figure compares total output in each year since 1970 with the output of those mines which existed in 1970, 1975, 1980 and 1985.

The output of each cohort of mines may decline as ore reserves become depleted, or it may increase as a result of new investment over the 20 year period. These opposing influences have balanced remarkably well. Those mines which were already producing in 1970 still produced almost 90% of their 1970 tonnage in 1990. Cyclical fluctuations, and temporary disruptions aside, the output of the subsequent cohorts was little changed in 1990 from their initial levels. This apparent stability has resulted from substantial investment to maintain output, which would have otherwise fallen dramatically. The chart brings out the extent to which 1990s' mine output was dominated by the mines

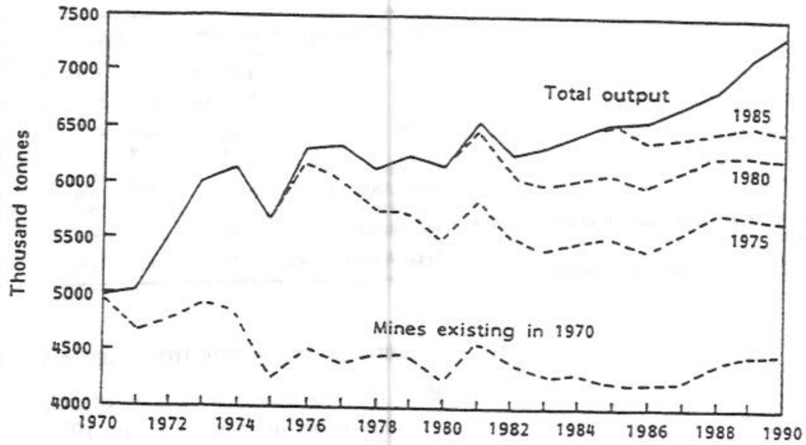


Figure 1. Western world copper output 1970–90: output from mines existing at different periods.

Source: RTZ Mine Information System.

which existed 20 years earlier, and conversely how increased demand was met largely from new mines. The distances between the layers, like the thickness of tree rings, reflects the industry's climate during each period, but with a lag. Thus the large rise between the lines for 1970 and 1975 mirrored the strength of copper markets in the early 1970s, while the small gap between the 1980 and 1985 lines picks up the recession of the early 1980s.

Similar analyses for each continent are illuminating, although continents do not closely coincide with politicoeconomic divisions. The lumpy nature of investment in new mines complicates any analysis in that the opening or closure of a large mine can upset apparent underlying trends. Nor do geological provinces follow geographical boundaries; one country or continent can embrace both openings and closures simultaneously. Table 1 shows how each continent's share of mine production changed between 1970 and 1990.

The third column of Table 1 shows shares of the 2.9 million tonnes of new capacity which came into production during the 20 years. Asia, Australia, Latin America, and more surprisingly, Europe, all attracted substantially greater shares of the new capacity than they had in 1970. Conversely Africa gained far less than its starting share.

The comparative attractions of the different continents changed considerably over the period, and they are still changing. Table 2 summarizes the changes in terms of each continent's contributions to changes in total output over each successive five-year period. The first column shows the changes in the output of existing mines over the relevant period and the second the part played by new mines. Here we can readily see the climatic differences changing. Use of 1975 as one terminal date may distort the first two periods because many US mines closed down in 1974–75 but reopened in 1976–78.

Table 1. Shares of Western world copper mine output 1970–90 (%).

	1970	1990	New capacity 1970–90[a]
Africa	26	16	7
Asia	6	8	14
Europe	4	5	8
North America	43	33	31
Latin America	18	31	30
Australasia	3	7	10
Total Western world	100	100	100

[a] Does not include expansions of mines that existed in 1970.

Source: RTZ Mine Information System.

Table 2. Changes in Western world copper output by quinquennia (thousand tonnes)

	1970–75		1975–80		1980–85		1985–90	
	Existing	New	Existing	New	Existing	New	Existing	New
Africa	–88	230	–121	140	–133	38	–182	27
Asia	–15	148	–15	62	–69	134	–4	4
Europe	26	58	–40	25	–5	59	–98	164
North America	–737	622	–152	92	–156	114	237	260
Latin America	–6	180	206	333	267	90	182	84
Australasia	81	192	–53	5	39	9	–186[a]	293
Total Western world	–739	1430	–175	657	–57	444	–51	832

[a] Primarily the enforced closure of the Bougainville mine.

Source: RTZ Mine Information System.

Some points arising from Figure 1 and Tables 1 and 2 are:

- The progressive erosion of existing African capacity, not matched by new investment in the 1980s.
- A turnaround in Latin America from the mid-1970s. Existing mines have expanded and there has been considerable investment in new facilities. The present five years will show a substantially greater rise than the 1985–90 period.
- Australasia has been in and out of favour, with substantial investment in the first and last periods. Low investment in the intervening years mainly reflects weak markets. The final period is distorted by the enforced closure of Bougainville Copper.
- Asia saw a dearth of investment in the last five years.
- By contrast, Europe did relatively well in the last period, mainly because of a large development in Portugal.
- North America displays the most interesting trend. Output declined up to 1985 with the production of new mines unable to offset closures of existing mines. The pattern was very similar to that of Africa. The last five years show a marked reversal, as the US industry made conscious attempts to regain competitiveness. Not only was there large investment in new capacity, but existing mines increased their output. The change has been so pronounced as to raise questions over the inevitability of any trend persisting. Thus the African decline, which has gathered momentum in the past two years, could perhaps be largely reversed by appropriate managerial and government policies.

Movements in relative mining costs

While the reserves of any individual ore body are ultimately finite, the precise timing of their exhaustion depends on economic conditions. The richness, accessibility and 'processibility' of ore deposits vary widely from one to another. How and why seemingly attractive ore bodies host relatively uncompetitive mines and, conversely, how highly efficient and competitive operations can be based on relatively indifferent resources are interesting questions, rather outside the scope of this paper. What is immediately relevant is that North America's declining output of copper in the 1970s and early 1980s described above reflected economic conditions, and, above all, deteriorating relative costs, mainly in the USA rather than Canada.

Both absolute levels of cost versus the world average, and changes in relative costs, are important indicators of competitiveness. The former reflect inherent geographical and geological features as well as domestic factor prices, exchange rates, taxes and institutional arrangements.

Table 3. Relative copper mine operating costs per tonne of ore milled (index numbers, Western world = 100; period averages in US$).

		1971–75	1976–80	1980–85	1986–89
Australia	Underground	142	132	166	148
Canada	Open pit	72	78	67	69
	Underground	122	143	153	162
Chile	Open pit	254	161	105	105
	Underground	99	42	35	37
Mexico	Open pit	85	67	94	64
	Underground	166	147	148	123
Papua New Guinea	Open pit	71	88	89	120
Peru	Open pit	116	88	96	109
	Underground	155	136	136	148
Philippines	Open pit	65	75	79	81
	Underground	23	20	34	33
South Africa	Open pit	70	72	79	87
	Underground	111	139	146	168
United States	Open pit	89	97	112	94
	Underground	62	75	68	72
Zaire	Underground	167	135	153	209
Zambia	Underground	90	102	78	90

Source: RTZ Mine Information System.

Differences can be enhanced or offset by differences in ore grades, and for some minerals, the values of associated products or byproduct credits. Variations in mining and milling costs per tonne of ore milled, which are dominated by domestic factor prices, can be counterbalanced by many of these specific differences. Thus Zaire's relatively high costs have been countered by copper grades which are substantially higher than average. The high costs of underground mining in Australia, Canada and South Africa have been more than offset in the market place by sizeable contributions from byproduct credits. A comprehensive discussion of the various influences on relative costs in the copper industry is contained elsewhere.[1] Subject to the important caveats that national averages can disguise marked variations between individual mines, and that averages over a period may mask sharp annual changes, movements in relative mining costs are summarized, for the copper industry, in Table 3.

The operating costs of mining and milling each tonne of ore in each country are compared with the Western world's average in index number form. Separate comparisons are made, where appropriate, for open pit and underground mines. Most of the leading copper producing countries are covered.

Movements in relative costs from one period to another affect, and possibly also reflect, changes in geographical shares of production. One influence is undoubtedly the performance of exchange rates, which can exacerbate or ameliorate movements in underlying relative costs, even between successive five-year periods. For example, the relative improvement of both Chile and Zambia in the early 1980s reflected exchange rate policies. The relationship between movement in costs and market shares is not all one way because one of the most widely exploited methods of cutting costs is to increase output. Fixed costs are thereby spread over much greater tonnages. Changes in Chile's costs, for example, partly reflected expanding production during the 1970s as the traumas of the late 1960s and early 1970s gave way to relative political and economic stability. Table 3 brings out a relative deterioration in US costs of open pit mining (the predominant method in the USA) during the 1970s and early 1980s, and a subsequent improvement. The different trend for underground mining suggests that these shifts reflected far more than changes in exchange rates.

---

[1] Phillip Crowson, 'Copper', in John Tilton, Hans Landsberg and Merton Peck, eds, *Competitiveness in Metals: The Impact of Public Policy*, Mining Journal Books, London, 1992.

US experience

The trend in relative US costs is illuminating when viewed in conjunction with the changes in North American production, highlighted in Table 2. The causal interplay between costs and production levels is complex, as already noted. The US copper industry's experiences have a wider relevance than as a case study in management. They show that seemingly irreversible trends can change direction.

By the early to mid-1980s the US copper mining industry was faced with the hard decision of greatly reducing its costs of all types or of facing a continuous erosion in its output. The failure of its attempts to gain protection against imports after the International Trade Commission's hearing of 1984 threw the industry back on its own resources. During the previous decades it had become inward looking and complacent, and seemingly oblivious to the rapid strides of the industry overseas. It believed that it faced subsidized state owned competitors who were not responsive to the profit motive, and it tended not to analyse its own strengths and weaknesses.

Several factors contributed to the US copper industry's renaissance. Changes in ownership and management swept away the traditions and shibboleths of the proceding decades, and forced a concentration on profitability. The brief incursion of the petroleum companies into the US mining industry paradoxically contributed greatly. Their entry into the copper industry was mainly based on false premises, as they soon discovered. For the pre-existing mining company managements, the clash of corporate cultures was, at best, extremely painful, and often terminal. The petroleum companies had mainly withdrawn within a decade, poorer and wiser, but they left a recapitalized and revitalized industry. Even where copper producers (like Phelps Dodge) were not assimilated by the oil companies, they were affected by this structural change.

Resurgent technical innovation was important: the most prominent example was dump and heap leaching, following by solvent extraction and electrowinning. This enabled the industry to exploit relatively high grade dumps and obtain a merchantable metal at a highly competitive cost. The copper ores of the south-western states of the USA are ideally suited to this technology. With this technology the dearth of byproduct values, which is a competitive disadvantage with conventional mining and milling, becomes a positive advantage.

Productivity improvements which had been pioneered elsewhere were exploited and extended. Having failed to modernize for so long, the US mines were able to make a step-wise jump when adequate capital was forthcoming. This often came from the new owners. The example of the Bingham Canyon mine may be extreme, but it does typify the entire US industry's experience. Kennecott, the owner, had been expropriated in Chile in the early 1970s and had also had some adverse experiences elsewhere. Its management concentrated on its interests outside the copper industry and Bingham Canyon became increasingly high cost. Kennecott was taken over by Sohio in the late 1970s, and that company was itself progressively dominated by British Petroleum. The Bingham Canyon mine was closed in 1983, and Kennecott was forced to dispose of most of its copper assets in order to support a modernization programme. Even that might not have been accepted had not Bingham Canyon's ore contained attractive gold credits. At a cost of some US$480 million, concentrating facilities were

113

135

upgraded and rationalized, methods of ore and waste haulage shifted from railways to diesel trucks, and the mine plan modified. The labour force of the whole operation was cut from over 7000 to barely 2000 when the mine reopened in 1988 and unit costs were slashed to the point where it has become one of the world's lowest cost producers of copper. The mine was acquired by the RTZ Corporation in 1989 and the modernization programme has continued, with investment in further milling facilities to allow reduced cut off grades and the utilization of ore that was previously treated as waste. Investment is now envisaged in expanded smelting and refining facilities. Given the necessary environmental and other permits, these will be operating by 1996 to allow further reductions in operating costs, and maintain Bingham Canyon's position among the world's lowest cost producers. They will incidentally bring substantial environmental benefits and set new standards in that field for the industry to attain.

The examples of Bingham Canyon and of the other success stories in the US copper industry in the late 1980s and early 1990s bring out several important messages for any mining industry.

- Complacency is dangerous and no management can afford to sit on its hands while producers elsewhere innovate and reduce their costs. No single producer or country has a monopoly of competence, and improved methods can come from any corner of the globe. In order to remain competitive management needs to copy and improve wherever possible.
- Salvation can only come from within the industry rather than from reliance on governmental action of any type. That is frequently reversed, with damaging consequences, or it is too little and too late. Nor is it feasible to rely on improved market conditions to bail out a mine in trouble. Prices do tend to move up cyclically, but seldom as or when expected. They also move down cyclically.
- In essence the task faced by mining companies' management is to achieve continuous improvements in relative costs and product quality against strengthening international competition in market conditions which are often adverse. More colloquially, it is akin to playing a game of leapfrog over competitors, while limbo dancing below a market driven price.

The importance of markets

The relative competitive stagnation of the US copper industry from the early 1970s until the mid-1980s sprang from its inward looking nature. It relied largely on domestic markets, and imports were mainly within integrated companies. A pronounced growth of imports of metal during the early 1980s was one of the main spurs to its change. The domestic industry was unable to supply adequate tonnages at the prevailing market prices.

Not all mining industries are able to meet the challenge of their domestic markets, particularly when these are expanding rapidly. Until the early 1960s, for example, Japan was largely self sufficient in many mineral products and metals but demand outstripped available supplies. Not only were domestic mines generally high cost, but there were insufficient local reserves. Japan effectively had two choices. It could build up its domestic metal industries on the back of imported ores and concentrates, or it could rely almost entirely on imported metals. In the

Table 4. Western world consumption of refined copper (thousand tonnes, annual averages).

	Western Europe	USA	Japan	Other countries	Total Western world
1958–60	1746	1229	223	380	3578
1963–65	2104	1686	412	556	4758
1968–70	2306	1830	774	615	5525
1973–75	2578	1871	970	854	6273
1978–80	2810	2077	1234	1178	7308
1983–85	2691	2088	1272	1335	7386
1988–90	3022	2188	1452	1891	8553
Increases					
1958–60 to 1968–70	560	601	551	235	1947
1968–70 to 1978–80	504	247	469	563	1783
1978–80 to 1988–90	212	111	209	713	1245
(%) increases from start of decade					
1958–60 to 1968–70	32	49	247	62	54
1968–70 to 1978–80	22	13	61	92	32
1978–80 to 1988–90	8	5	17	61	17

Sources: Metal statistics (successive issues), Metallgesellschaft AG, Frankfurt am Main.

main, it followed the former route, at least up to the oil price rises of the 1970s. Then high domestic energy costs precipitated a complete run down of the sizeable Japanese aluminium smelting industry. Given the subsequent behaviour of energy prices, many Japanese now question the wisdom of that decision. In other metals, including steel, Japanese metal producers had to foster the rapid development of new sources of raw materials in competition with the well established trading links of North America and Western Europe. Other countries, notably in South-east Asia, have later followed a similar route to the Japanese.

Table 4 shows changes in the location of demand for refined copper over the past 30 years.

Three yearly averages are used to minimize the distorting effects of cyclical fluctuations in demand. The relatively rapid growth first of Japanese and then of 'other countries' demand from the mid-1960s onwards is clearly brought out.

The mines developed to supply the new sources of demand of the last 30 years naturally incorporated the latest available techniques, and were often much larger in scale from start up than the existing mines. That gave them the additional advantage of economies of scale. Designed from the outset to supply export markets, they were able to develop dedicated transport and handling systems. They benefited in turn from the substantial reductions in shipping costs of the 1960s and early 1970s which were mainly associated with increased vessel sizes.

We have already noted that the easiest way to reduce costs is to spread their fixed components over increased tonnages. Hence mines facing rapidly rising demands can more easily respond to competitive circumstances than those faced with static or declining offtake.

Rising capital costs made it increasingly difficult to finance new projects with equity funds, particularly against a background of exchange rate instability and varying political risks. Project finance via debt became the main route for developing new mines, and this has generally required the backing of sufficient long-term sales contracts as security for payments both of interest and principal. Gold has been a significant exception in this regard. Burgeoning Japanese demand, sometimes supplemented by European offtake, was the main source of such contracts. Interestingly, uranium mines were financed by a broader geographical spread of contracts because none existed until the 1940s.

The voracious Japanese appetite for additional mineral supplies of the 1960s and 1970s inevitably encouraged exploration and mineral development in new areas. The resource bases of the existing suppliers were in most instances unable to support adequate expansion to meet demand. Even with improved ocean transport, many of the mines clustered around the Pacific rim rather than the Atlantic ocean. The pace of Japan's growth of demand has eased back, and other countries have taken over its spearheading role. Other South-east Asian countries have been in the forefront, and have provided supplementary markets to the Japanese. In some instances, they too have sought ores and concentrates rather than primary metals, as they have followed the Japanese route to development.

As the initial Japanese contracts matured and, in many cases before then, the new mines sought to diversify their markets to reduce their dependence on the Japanese. There are unsubstantiated suspicions among mineral producers that the Japanese industry exaggerated its needs in order to encourage excess capacity and keep prices in check. Such suspicions undoubtedly benefit from hindsight, but may have some substance. In any event, mines have diversified from their original markets and intensified competitive pressures on traditional suppliers to the more mature markets.

Rising Japanese and Asian demand has also affected mine location and competitive pressures indirectly via user industries. At the extremes, Japanese competition has forced the closure of European or US manufacturing capacity with knock on effects back up the chain on the original raw material producers. In other instances strong price competition has forced the metal and mineral industry's customers to seek low cost sources of supply. In the non-ferrous metals, which are priced largely on terminal markets, this competitive pressure has been no less real, if more diffuse, than in those mineral industries which have more inflexible pricing systems.

Individual experiences

Moving away from copper, the experiences of other minerals both reinforce the points already made and bring out other aspects of location in order to form a more balanced view. The starting point for the discussion of each product is how shares of world production have altered over the past 30 years.

Bauxite

Total output of bauxite rose from 27.7 million tonnes in 1960 to 60.7 million tonnes in 1970, 92.7 million tonnes in 1980 and 112.7 million tonnes in 1990. There were cyclical peaks and troughs within the period. Table 5 brings out some of the main features of changing market shares, which were dominated by the failures and miscalculations of government policy in the Caribbean producing countries, and the rise of new producers, helped by falling transport costs and the changing location of markets.

Absolute declines in French and US production reflected reserve exhaustion and rising extraction costs. Stagnation and then decline in the USSR over the past twenty years mirrored domestic developments in demand.

Production peaked in the Carribean countries in the early 1970s and

Table 5. Shares of world bauxite production (%).

	1960	1970	1980	1990
Australia	0.3	15.3	29.3	36.7
Brazil	0.4	0.8	4.5	8.8
France	7.5	5.0	2.0	0.4
Guinea	5.0	4.1	15.0	15.5
Guyana	9.1	7.3	3.3	1.3
Jamaica	21.1	19.8	12.9	9.7
Surinam	12.5	9.9	5.3	2.9
USA	7.3	3.5	1.7	0.4.
USSR	12.9	10.7	6.9	4.7
Other countries	23.9	23.7	19.1	19.6
Total	100	100	100	100

Sources: US Bureau of Mines, *Minerals Yearbook 1964*, Government Printing Office, Washington, DC, 1965 and *Metal Statistics*, Metallgesellschaft AG (successive issues).

subsequently fell following nationalizations. These countries, led by Jamaica, badly miscalculated their ability to impose taxes and raise prices substantially. Their attempts to do this coincided with, and also accelerated, the discovery and development of new ore deposits, notably in Australia, Brazil and Guinea. Once customers had switched to the new ores, sometimes only after modifying their plants, they were reluctant to go back to their traditional suppliers. When those came to repeal their earlier fiscal mistakes, demand was no longer growing to absorb additional output. Their experiences emphasize that possession of large resources and even the existence of production facilities are insufficient to maintain market shares.

The development of the newer sources of bauxite was facilitated by the spread of alumina and aluminium production away from the USA and Western Europe, and by a lowering of shipping freight rates. Australia, Brazil and Guinea raised their output by 56.5 million tonnes between 1970 and 1990 against an overall world increase of 52 million tonnes.

Gold

World gold production grew from some 38 million ounces in 1960 to 47.5 million ounces in 1970, but fell back to 39 million ounces in 1980. It rose sharply in the 1980s to reach almost 65 million ounces in 1990. The trend in total output, and the movements in countries' shares, emphasize the driving influence of expected profits on exploration, and the willingness of the mining industry to extend its geographical frontiers where the rewards are expected to outweigh the risks. Furthermore the experiences of the 1980s show that mining does not necessarily migrate away from the mature industrial nations of the OECD. Table 6 gives shares of world gold production from 1960–90.

During the 1960s when prices were pegged at $35/ounce, production became unprofitable and declined in most countries, except South Africa and the USSR. The former benefited from a favourable tax regime.

Rising prices in the 1970s, culminating in the 1980s' spike, greatly

Table 6. Shares of world gold production (%).

	1960	1970	1980	1990
Australia	2.8	1.3	1.4	12.0
Canada	12.1	5.1	4.2	8.2
South Africa	56.0	67.7	55.4	30.0
USA	4.4	3.7	2.5	14.6
USSR	10.7	13.7	21.2	8.7
Other countries	14.0	8.5	15.3	26.5
Total	100	100	100	100

Sources: USBM *Minerals Yearbooks* (successive issues) and Consolidated Gold Fields/Gold Fields Mineral Services, *Annual Reports* on gold.

Table 7. Shares of world iron ore production (% based on iron content).

	1960	1970	1980	1990
Australia	0.9	7.3	11.6	12.4
Brazil	2.0	5.9	14.4	17.6
Canada	4.1	6.7	5.9	4.0
China	9.1	4.9	7.2	15.7
France	6.9	4.0	1.7	0.5
India	3.4	4.4	4.9	5.5
Sweden	4.7	4.6	3.4	2.3
USA	18.8	12.9	8.6	6.1
USSR	19.3	24.2	26.0	22.8
Other countries	30.8	25.1	16.3	13.1
Total	100	100	100	100

Sources: USBM *Minerals Yearbooks* (successive issues) and International Iron and Steel Institute's *Steel Statistics*, 1991.

encouraged exploration, which initially focused on old mining areas. Improved technology, partly developed by the US Bureau of Mines, facilitated the treatment of old mine dumps and previously unattractive deposits. Exploration was initially concentrated in economically and politically stable regions. In Australia and Canada, it was greatly assisted by favourable tax treatment. Production rose strongly in these two countries and the USA in the 1980s. It also grew considerably in Europe and Japan, although their shares of total output are low.

South African output has in the meanwhile declined, caught by a cost/price squeeze and political uncertainties.

In the mid- to late 1980s exploration interest shifted to the volcanic zones of the Pacific rim, and even to the older gold producing areas of Africa. Their output is now rising, but political factors and weakened gold prices are restraining investment.

Iron ore

World production of iron ore on a gross weight basis grew from roughly 520 million tonnes in 1960 to nearly 770 million tonnes in 1970, almost 900 million tonnes in 1980 and just under 1 billion tonnes in 1990. The geographical patterns of growth over the last 30 years reinforce the earlier discussion about the roles of growing market demands and transport costs. Rising Japanese demand was a strong motive force. Production of iron ore is geographically more diffuse than of some other minerals. Shares of major producing countries are summarized in Table 7.

The USSR has been mainly self sufficient. Also the US industry has been closely linked to US steel production. Its relative decline mainly reflects the performance of its downstream markets. US iron ore output was flat in the 1960s but has subsequently fallen.

Chinese output has grown but not sufficiently to meet local needs, and imports of higher grade materials have increased, especially from Australia.

French output has dropped, partly as domestic ores have been exhausted, but mainly because falling shipping costs have made higher grade imported ores from Brazil, Africa, and even the Pacific region relatively more attractive than the much lower grade European ores.

Hidden within the other countries' shares, those of Latin American producers, such as Chile, Peru and Venezuela, declined in the 1970s following nationalizations, the severing of established marketing links and subsequent underinvestment. Liberia's share has collapsed through a combination of reserve exhaustion and civil war. Conversely, South Africa's share has risen in spite of the 1980s' sanctions because of aggressive pricing and strong governmental support for rail and port developments.

Table 8. Shares of world nickel mine production (%).

	1960	1970	1980	1990
Australia	–	4.5	10.1	7.7
Canada	56.9	41.8	25.0	22.8
Cuba	4.2	5.5	5.2	4.9
Dominican Republic	–	–	2.4	3.3
Indonesia	0.1	1.6	5.5	6.2
New Caledonia	15.7	20.9	11.7	9.7
Philippines	–	–	4.7	1.8
USA	3.3	2.1	1.4	–
USSR	17.0	16.6	19.4	24.2
Other countries	2.8	7.0	14.6	19.4
Total	100	100	100	100

Sources: Metal Statistics (successive issues), Metallgesellschaft, Frankfurt am Main.

Sweden's production and shares have fallen in the face of the slow growth of its European markets and competition from imported ores, but not as much as might have been predicted because of the development of niche markets. LKAB, the Swedish producer, has concentrated on improving product quality, and meeting specific customer requirements, especially for pellets.

The main gainers have been the rapidly expanding industries of Australia and Brazil. Steel mills in the two regional markets of Europe and East Asia have, none the less, encouraged other suppliers, and have played the leading producers off against each other. In Brazil the centre of gravity is shifting north to the Amazonian region following the internationally supported development of the Carajas railroad.

Nickel

World production of nickel has become much more geographically diversified over the last three decades. One of the motive forces was a jump in prices in the late 1960s caused by a Canadian strike. Physical shortages encouraged exploration, which broadened the geographical spread of production. Production expanded from 342 000 tonnes in 1960 to 663 000 tonnes in 1970. Growth slowed in the next decade, as it did for most minerals and metals, but reaccelerated during the 1980s. By 1990 world production had reached 874 000 tonnes. From 1970 onwards movements in relative costs, especially of energy, have been important influences. The percentage shares of production have developed as shown in Table 8.

The exploration boom triggered by the late 1960s' shortages led to discoveries in Australia, and also the development of lateritic ores in tropical regions. The Canadian producers themselves diversified overseas, for example in the Dominican Republic and Indonesia.

Canadian output grew during the 1960s but dropped in the next decade and was expected to drop even further. The tropical lateritic ores are, however, very energy intensive. The jumps in oil and energy prices of the 1970s markedly altered the relative economics of lateritic and sulphide ores as mined in Canada. The latter were given a new lease of life, and are now among the lowest cost producers. By contrast, the energy intensive New Caledonian output declined, with political uncertainties there giving a further jolt. Similarly, US production was forced to close by its high costs (although it has since restarted on a very small scale). Relatively new mines were closed in Guatemala and the Philippines, and the Dominican mine was also restricted for some years.

Russian and Cuban output grew, partly to meet defence needs, but also to provide a valued source of foreign exchange.

One important influence on Canadian production, and a cause of

Table 9. Shares of world phosphate rock production (%).

	1960	1970	1980	1990
China	1.5	2.0	4.0	11.2
Christmas Island	1.2	1.3	1.2	–
Israel and Jordan	1.4	2.4	4.5	6.1
Morocco	17.9	13.4	13.6	13.9
Nauru	3.3	2.5	1.5	0.6
Tunisia	5.0	3.6	3.3	4.1
USA	42.5	41.4	39.3	30.1
USSR	16.7	24.5	21.3	21.7
Other countries	10.5	8.9	11.3	12.3
Total	100	100	100	100

Sources: USBM *Minerals Yearbooks*.

reduced output in the 1980s, was concern about the environmental damage of sulphur emissions from local smelting. This forced substantial abatement costs on the producers and also entailed highly selective mining. The lateritic ores, which generally produce a Class II rather than Class I nickel, do not have the same environmental burdens.

Phosphate rock

Production of fertilizers is the main outlet for phosphate rock, and this is more dependent on agricultural needs than on the industrial production which drives demand for most minerals. Production, therefore, did not suffer the same set backs in the 1970s as other products, but it suffered in the 1980s from the debt constrained problems of many importing developing countries. Total world output doubled from 42 million tonnes in 1960 to 85 million tonnes in 1970. It had grown to 138 million tonnes by 1980, and 154 million tonnes by 1990. The changing shares of the leading producers, which are summarized in Table 9, reflect the exhaustion of some producing regions, and the impact of government and commercial policies. Consumers fought to diversify away from the dominant producers, who, as in bauxite, overestimated their market power.

The production of Christmas Island and Nauru has dwindled from the 1970s' level because of reserve exhaustion, the ultimate reason for loss of competitiveness.

The US producers, with an official export cartel, and Morocco, have dominated world trade. Morocco attempted to boost prices substantially in the early 1970s, to be faced by a consumers' strike and the development of production elsewhere. Israel and Jordan were two beneficiaries. Even so, Moroccan output has risen and the share has remained fairly stable.

US output rose strongly, if not quite as fast as the total, until the early 1980s. Then the existing mines began to run down, but access to new ore zones ran into major environmental obstacles, especially in Florida, the main producing state. Production fell in the 1980s, and further declines are inevitable, although abundant resources remain. This is one clear instance of environmental requirements influencing competitiveness. Other countries' phosphate deposits are not located in regions anywhere near as ecologically sensitive as the Florida wetlands.

Zinc

For a relatively mature mineral, which is mined predominantly in the OECD countries, there have been some substantial changes in the location of mine output. World output increased from 3.4 million tonnes in 1960 to 5.6 million tonnes in 1970, 6.3 million tonnes in 1980 and 7.3 million tonnes by 1990. The OECD countries have remained dominant

WHARTON REPROGRAPHICS

LIPPINCOTT LIBRARY
DOCUMENT DELIVERY

How Domestic Demand Shapes the Pattern of International Trade

Rolf Weder

1. INTRODUCTION

THE relationship between domestic demand and the pattern of trade has not received great attention in international trade theory so far. This has been observed by Hunter (1991) and Melvin (1993) among others. The situation contrasts with the seminal work by Linder (1961), emphasising precisely this relationship, and with a stream of recurring empirical evidence suggesting that differences in demand patterns among countries and regions seem to be important determinants of the pattern of trade.

This paper attempts (i) to identify the main demand characteristics which are likely to influence the pattern of trade among countries and (ii) to present some underlying reasons for differences in demand conditions among countries. The adopted methodology is to develop a taxonomy of demand characteristics and to illustrate it with empirical evidence from recent case study analyses. The taxonomy itself is based on the essence of a few models emphasising demand and trade — which nevertheless have evolved in international trade theory — as well as on a condensation of the many facets of the demand condition described in Porter (1990) and others.

It should be noted that traditional trade theory, i.e., the classical and neoclassical theory of comparative advantage, is not in contradiction with observations suggesting that domestic demand creates conditions under which domestic firms develop expertise in certain industries and, in turn, establish a comparative advantage of the country. The pattern of trade remains determined

---

ROLF WEDER is from the Department of Economics, University of Basel. A first draft was written while he was a postdoctoral fellow at Simon Fraser University. He would like to thank the members of the Department of Economics for providing such a stimulating environment, and Silvio Borner, Steve Globerman and Herbert Grubel for their comments on earlier drafts. A number of very helpful suggestions by two anonymous referees have greatly improved the paper. Financial support from the Swiss National Science Foundation, Grant No. 8210 030613, and from the WWZ-Förderverein are gratefully acknowledged.

© Blackwell Publishers Ltd 1996, 108 Cowley Road, Oxford OX4 1JF, UK
and 238 Main Street, Cambridge, MA 02142, USA.

by comparative advantage as advocated by this theory. However, traditional theory cannot *explain* the relationship between domestic demand and the pattern of trade as long as comparative advantage is exogenous. Models of 'new' trade theory based on imperfect competition thus appear to provide a better basis for an adequate analysis of the role of domestic demand in shaping the pattern of trade (Weder, 1995a).

The remainder of the paper is as follows. Section 2 proposes a taxonomy of demand characteristics which are likely to influence the pattern of trade on the basis of theoretical models and empirical observations. Section 3 describes some case study evidence to support this suggestion. In Section 4, determinants of observed country-specific demand characteristics are discussed. Conclusions are found in Section 5.

2. TOWARDS A TAXONOMY OF DEMAND CHARACTERISTICS AND TRADE

Linder (1961, p. 87) argued that countries would export those products for which there exists a large domestic market or, as he called it, a 'representative demand'. According to him, this effect would arise because firms tend to become aware more easily of business opportunities at home, because invention and innovation is more likely inspired by unsolved problems in the home market and because the continuous adaptation and improvement of a new product incurs significantly lower costs if undertaken for buyers in the home market. This mechanism has been adopted by Vernon (1966) in his international product cycle approach, and further facets of home demand and their positive impact on exports are also described by Porter (1990, p. 86) who, in addition, discusses features like the growth of demand, the change of demand structures or the number of buyers.

The taxonomy established below attempts to boil down the complex picture found in studies by Porter and others. As shown in Table 1, it will be distinguished between 'composition' and 'sophistication' of domestic demand, on the one hand, and between 'static' and 'dynamic' influences from domestic demand, on the other. Note that domestic demand not only includes final consumer goods and services (e.g., clothing or banking services for private customers), but also derived demand for intermediate products and services (e.g., textile dyestuffs or transportation of bank notes).

The following discusses how these characteristics are likely to affect the pattern of trade. The discussion takes into account a few models in this field which have emerged in international trade theory. I first describe static and dynamic aspects of the composition of domestic demand, then switch to the sophistication and conclude.

© Blackwell Publishers Ltd 1996

TABLE 1
Taxonomy of Key Characteristics of Domestic Demand

Key Characteristics	*Static*	*Dynamic*
Composition	Horizontal Segment Structure	Growth and Internationalisation of Segment Structure
Sophistication	Vertical Segment Structure	Learning and Competition among Buyers

a. Composition

The *horizontal segment structure* of domestic demand describes the relative size of demand for different varieties in an industry. There exist two approaches in international trade theory creating a positive relationship between relative home market size and exports of a product. First, assuming an explicit movement from autarky to free trade and some adjustment costs, countries tend to have a relative cost advantage in those products where domestic demand is relatively larger if the production is characterised by increasing returns to scale or if it requires R&D investment in product innovation. Second, the same result emerges if international transportation or, more generally, transaction costs are combined with increasing returns to scale providing firms in large markets with a home market advantage.[1]

The dynamic aspects — high *growth* of domestic demand for certain varieties or an *internationalisation* of the segment structure — may help and reinforce the creation of competitive advantage by firms as observed by Porter (1990, p. 94). First, it is suggested that a high growth rate of domestic demand leads to a faster adoption of new technologies by firms and to a quicker replacement of old inefficient plants. Firms may thus develop a competitive advantage in product segments which are characterised by this home demand effect. Second, an internationalisation of domestic demand patterns to foreign markets is likely to reinforce the home country's comparative advantage if adjustment costs and learning effects in production are involved. Foreign competitors are, in a way, misled by their domestic demand, whereas the home country's firms benefit from a demand pattern which anticipates world-wide trends.

b. Sophistication

The *vertical segment structure* of domestic demand describes the relative size of demand for products of a different quality in an industry. In contrast to the

---

[1] Models of the first group include Grubel and Lloyd (1975) — based on increasing returns to scale — and the 'biological' models by Bhagwati (1982) and Dinopoulos (1988) — based on R&D. The second group of models has been established by Krugman (1980). Weder (1995b) generalises and interprets this latter model.

© Blackwell Publishers Ltd 1996

horizontal segment structure, there is a clear hierarchy among vertically differentiated products. Consumers always prefer a higher quality to a lower quality variety. As higher quality products require more effort and resources in their production, they are more costly and thus have a higher price. Consumption being income-constrained, the average consumer with a lower income will choose lower quality products.[2] Thus, differences in domestic demand among countries for vertically differentiated products are likely to arise from differences in per capita income or income distribution. Based on the positive relationship between sophistication of domestic demand and exports of quality differentiated products suggested by Linder (1961), one would generally expect rich countries to export higher quality products and import those with a lower quality.

This is, in fact, what the international trade models by Falvey and Kierzkowski (1987) and by Flam and Helpman (1987) propose. The first approach is rooted in a Heckscher-Ohlin setting where countries differ in their capital-labour endowment — and thus in per capita income — and where the production of higher quality products requires a greater capital-labour input ratio. It turns out that the relatively capital-abundant country exports the higher quality products importing lower qualities which is in line with the Heckscher-Ohlin Theorem. The second approach by Flam and Helpman is based on the Ricardian model where the richer country is assumed to have a comparative productivity advantage in high quality products. Again, the richer country tends to consume and to export higher quality products whereas it imports the lower quality differentiated products.

For the purpose of this paper, it is important to note that the positive relationship between domestic demand for and exports of high quality products is directly based on the assumed factor endowment or technology in both models. For example, in the Falvey-Kierzkowski approach, the relatively capital-abundant country exports the relatively high quality products because capital is the factor intensively used in the production of high quality products. At the same time, however, the capital-abundant country's domestic demand is biased towards high quality products because of the higher per capita income resulting from the factor endowment. Thus, the underlying endowment simultaneously causes specialisation in both consumption and production.

But there is no effect from domestic demand on comparative advantage. Rather to the contrary. In both models, the preference for high quality products in the richer country, in principle, creates an incentive to import the preferred quality which, however, does not 'break through' because the richer country has an assumed comparative advantage in high quality products. Thus, the two approaches produce a result where both production and consumption of high

---

[2] See also Greenaway (1987, p. 99) for a discussion of different models in this field.

© Blackwell Publishers Ltd 1996

quality goods are positively correlated without either causing the other.[3] They cannot capture the demand effects considered in this paper where domestic demand *causes* a comparative advantage in the production of certain goods.

Two dynamic aspects of the sophistication of domestic demand are effects from *learning* of domestic buyers and *competition* among them which, according to Rosenberg (1982, p. 120) and von Hippel (1988), may further increase the degree of sophistication over time. As buyers learn by using a product, they are important sources of information for innovators and become demanding potential customers of local suppliers. Similarly, competing buyers may develop tough requirements for input producing firms in order to remain competitive. Note that competition among buyers of consumer goods may result from a sociologically determined conspicuous consumption where each buyer extremely values the distinction between his or her product and similar products owned by other people — a point which has already been described by Veblen (1899, p. 68) in his analysis of the behaviour of the 'leisure class'.

c. *In Sum*

This short discussion leads to the conclusion that international trade literature has given some attention to the effects from the horizontal domestic demand structure on the pattern of trade. These models assume different tastes among countries and study trade in final consumption goods. However, the impact from the vertical demand structure and, in particular, from the dynamic aspects of demand characteristics on the export performance have not been studied carefully so far. Furthermore, the effects from differences in derived demand for intermediate goods seem to be rather neglected and the underlying reasons for differences in demand among countries are more or less assumed to arise from differences in preferences.

This provides the basis for the remaining two sections which give some examples for the simple taxonomy presented above and the underlying reasons for differences in demand among countries. The target is also to indicate how the observed relationship between demand and trade could be looked at in a more general, theoretical way.

---

[3] This important distinction between correlation and causation has been brought to my attention by one of the referees. Note that the same kind of result is created in Markusen (1986) and Hunter and Markusen (1988) who assume a non-homothetic utility function in a Heckscher-Ohlin setting. Their model implies that the trade volume among countries with different factor endowments is smaller than predicted by a standard factor-endowment model with homothetic preferences because a relatively large demand for X creates a tendency to import X.

© Blackwell Publishers Ltd 1996

3. SOME EMPIRICAL EVIDENCE ABOUT DEMAND AND TRADE

The following examples are used to illustrate the relevance of the taxonomy proposed above. The description is based on industry case studies with all their well-known methodological advantages and disadvantages.

a. Demand Composition and the Pattern of Trade

Examples which provide a general support of a positive relationship between relative home market size for and exports of horizontally differentiated products include United States' commercial television shows and equipment (Porter, 1990, p. 300), Italian ceramic tiles (Porter, 1990, p. 214), Swedish Ikea-type furniture (Grubel and Lloyd, 1975, p. 97), Canadian equipment for resource-based industries such as forestry machinery (Globerman, 1990, p. 269) and Great Britain's leading position in exports of auctioneering services (Porter, 1990, p. 271).

A careful analysis of individual industries shows that firms typically have gained a strong and sustained competitive advantage in those segments where dynamic effects from domestic demand reinforced the impact from the horizontal segment structure. For example, the Swiss paper product manufacturing machinery industry was an early developer of labour-saving post-press machines for the Swiss and the nearby South German printing and newspaper firms in the 1960s (Borner et al., 1991, pp. 171). Domestic demand for this type of machine was relatively large in Switzerland and Germany because of relatively high wages, labour shortages and no objections by unions against improving labour productivity. This situation contrasted with the one existing in other countries such as the United Kingdom or the United States where the printing industry was heavily unionised and thus hostile to innovations. It is important to note that the Swiss and German firms in the supplying machinery industry achieved a world-wide leading position when their domestic demand pattern began to internationalise because of other countries' tendency to deregulate and to de-unionise the printing industries.

Dynamic aspects from demand were also important for the Japanese facsimile industry which benefited from a fast growth of domestic demand for facsimile machines (Enright, 1992, p. 1) as well as for the Swiss sugar confectionery industry which enjoyed the advantage from a relatively large and internationalising domestic demand for sugar-free herbal cough drops (Richterich, 1995, p. 343).

b. Demand Sophistication and the Pattern of Trade

A positive relationship between the relative home market size of vertically differentiated product segments and exports is suggested by the Japanese

© Blackwell Publishers Ltd 1996

electronics industries (e.g., cameras and stereos) and the Italian clothing and footwear industries.[4] It is argued that the Japanese have been passionate buyers of consumer electronics products with a desire to adopt the latest model because these products were important status symbols. Similarly, Italians have traditionally been sensitive to new styles and fashions in areas like clothing and footwear and they quickly adopted new designs in order to differentiate themselves from each other. Another example is Sweden where an extreme concern for safety and environmental quality, reflected in domestic behaviour and regulations, helped to create a car industry which established a world-wide reputation for safe cars (e.g., Volvo).

These and other examples indicate that the relationship between demand and trade is caused or at least reinforced by domestic demand characteristics and thus not a pure outcome of, e.g., differences in factor endowment as discussed in Section 2*b*. Moreover, the joint role of static and dynamic aspects seems to be important. Countries have established a comparative advantage in those quality segments where domestic demand was relatively large and where dynamic effects from competition and learning among buyers was relatively strong — be it because of a conspicuous behaviour among individuals to differentiate themselves from others, or because of a considerable pressure on indigenous firms to produce high quality finished products and, thus, to purchase high quality intermediate inputs.

However, note that the vertical segment structure of home demand may also provide a disadvantage to domestic firms if, e.g., demand for a high quality and thus expensive product is large at home, but non-existent or declining abroad. Such a situation may have been an important disadvantage to the shrinking Swiss textile machinery industry of the 1980s. This industry has been facing an extremely luxury-oriented domestic textile industry which became less representative of foreign demand (Bandelier, 1995, p. 278).

4. DETERMINANTS OF DIFFERENCES IN DOMESTIC DEMAND

The preceding analysis showed that differences in domestic demand for final and intermediate products among industrialised countries exist and that they affect the pattern of trade. These differences just have been taken as given for a certain industry and at a certain point in history. But where do these differences in domestic demand come from? Are we not walking on thin ice by emphasising demand as an important determinant of trade patterns, given the fact that most models in international trade theory assume identical preferences throughout the

---

[4] These examples are taken from Porter (1990, pp. 346, 404 and 440).

© Blackwell Publishers Ltd 1996

world? Stigler and Becker (1977, p. 89) even state: '(...) no significant behavior has been illuminated by assumptions of differences in tastes'.

In this section, we want to look more carefully at the sources of demand differences. The examples are mainly taken from Swiss industry analyses (Enright and Weder, 1995), even though similar evidence is likely to occur in many other countries and industries.

a. Income

The income of individuals can be expected to influence the share of expenditure spent on segments of horizontally or vertically differentiated products. Thus, differences in per capita income and in the income distribution among countries may be important sources of differences in the composition and sophistication of domestic demand.

An interesting example is provided by Switzerland whose high level of per capita income, among the highest since the first world war, has created a high ability to pay for a relatively large group of citizens. This, in turn, established a relatively large and early domestic demand for high quality products in internationally successful Swiss industries such as watches, chocolate or textiles or for healthy products such as sugar-free herbal cough drops. Wealthy tourists and business people from abroad may have amplified the sophistication of domestic demand in some industries.[5] Moreover, the particularity of the Swiss consumer goods industries also created a corresponding pattern of derived demand for high quality intermediate products such as textile dyestuffs, textile machinery, comfort controls and fire detection equipment.[6]

The implicit model which may be able to capture this situation is a combination of approaches discussed in Section 2. A non-homothetic utility function would allow for different patterns of demand for vertically and horizontally differentiated products among countries with different levels and distributions of income. Increasing returns to scale and transaction costs, e.g., would create a tendency towards the export of the relatively preferred goods. In order to get beyond the described Falvey-Kierzkowski approach where capital-abundant Switzerland would export and consume high quality goods without either causing the other, one could, for example, assume a positive impact from demand-induced production of high quality products on capital accumulation. This may create a cumulative effect for a country with a slightly higher income or a less equal income distribution.

---

[5] Fifty per cent of Swiss domestic demand for watches appear to stem from foreign visitors (Bumbacher, 1995, p. 143).

[6] These cases are described in Weder (1995c), Bandelier (1995) and Enright (1995a and 1995b), respectively.

© Blackwell Publishers Ltd 1996

b. Total Relative Prices

Relative prices are important determinants of the composition of domestic demand for final and intermediate products. However, it can be expected that agents choosing the optimal consumption or input bundle do not only take into account the price of good *A* relative to *B* itself, but also the price of the complementary goods *C* or *D* associated with the use of good *A*. Thus, they consider the 'total' or 'full' relative price of *A*.[7]

An interesting example on the intermediate product level is the Swiss paper product manufacturing industry described above which has been successful in inventing and exporting labour-saving machines for firms printing newspapers and magazines. Demand of the domestic Swiss printing plants was relatively large for machines embodying a labour-saving technique — not because these machines themselves were relatively cheap on the domestic market, but rather because the *total price* including the costs to operate these machines was relatively lower, given the high Swiss labour costs. Another example of this type is the relatively large Swiss share of expenditure on sugar-free confectionery products.[8] Suppose again that consumers of sugar confectioneries take into account any additional costs which are possibly associated with the consumption of these products such as dental treatment. The pattern of Swiss domestic demand is thus in line with the fact that dental insurance is not part of the government supported health care system and relatively expensive.

The described situation on the intermediate product level implies a model which creates demand effects as discussed in Section 2*a* where, for example, firms invent those horizontally differentiated intermediate products which face a relatively large domestic demand. Capital-abundant countries would then export those durable products which are not only capital-intensive in production, but which also are capital-intensive in use and thus embody a capital-intensive technique with respect to their application.

c. Government Regulations

The composition and sophistication of domestic demand is also influenced by the government who may either be an important buyer of goods and services in some industries (public procurement) or who may regulate private demand (taxes, standards).

It appears to be rather a disadvantage for firms and industries if sales to the government represented a large share of their total business — at least in the

---

[7] This is an adaptation from the address models of product differentiation; see Eaton and Lipsey (1989).

[8] In 1993, e.g., this share was 21 per cent of total Swiss demand for sugar confectionery products — the highest share among all industrial countries; see Richterich (1995, p. 343).

© Blackwell Publishers Ltd 1996

Swiss case. Examples include truck manufacturing and telecommunications where Swiss suppliers have indirectly been subsidised through public procurement and thus not been able to become internationally successful.

Examples of the impact from government regulations are given by the fire detection equipment and the dyestuff industries. Swiss fire detection codes have been established early in history and were among the most exacting and demanding in the world. This early regulation may have created pressure on the fire detection industry to cope with high standards at the origin of the industry. Most likely it contributed to the emergence of a relatively large Swiss market for fire detection equipment (Enright, 1995b, p. 337). Similarly, the early and demanding Swiss water effluent regulations made the indigenous textile industry — especially the dye mills — seek textile dyestuffs with fewer problems in water effluent treatment. Not surprisingly, Swiss dyestuff producers have attained a strong international position in this product segment (Weder, 1995c, p. 48).

To the extent that government regulations are the result of a political process that allows the complete electorate to voice tastes, they are part of consumers' preference system and thus could simply be reflected by the utility function in a model of the type described in Section 2*a*. For example, Swiss emphasis on fire protection may ultimately result from Swiss tastes for fire protection that find their way into established laws.[9] However, one may also argue that regulations are the result of rent-seeking by special interest groups which then would suggest that regulations should be modelled as constraints on the individual agent's maximising behaviour.

d. Climate and Topography

Climate and topography seem to be more important determinants of the patterns of domestic demand than expected — at least from a historical perspective. As these sources of differences in demand among countries hardly change, they might easily be integrated directly in the preference system of individuals or in production functions of firms.

For example, the variations in the Swiss *climate* — between winter and summer, but also within the seasons — created a domestic demand for powerful, reliable and flexible comfort control systems which has been reinforced by the high per capita income (Enright, 1995a, p. 298). Most interestingly for the purpose of this paper, the lacking necessity to air-condition houses and office buildings during summer did not create a large demand for air-conditioning equipment — a product segment in which the Swiss comfort controls industry never attained a strong international position.

---

[9] This point has been brought to my attention by one of the referees.

© Blackwell Publishers Ltd 1996

The special Swiss *topography* — from extremely steep and narrow areas in the mountains to open and flat surfaces in the centre of the country — created a special domestic demand for transport equipment, road construction, bridges and houses. Swiss trucks, for example, have been known to be powerful, extremely durable, easy to handle on windy roads and deliverable in all length and width. They perfectly matched the requirements of a relatively large domestic demand. Even though Swiss truck manufacturers (Berna, Saurer) have obviously been influenced by domestic demand characteristics, this composition of domestic demand has not been proved to be an advantage in the long-term because of other reasons — including the possibility that Swiss demand patterns did not internationalise.[10]

Another example is the Swiss freight forwarding industry. The particular Swiss topographic situation created a large 'domestic' demand for co-ordination and handling services for foreign firms as goods to be transported through Switzerland had to be unloaded from various foreign transport systems to particular domestic ones at different times in history. The Swiss freight forwarders exploited this experience from the home market internationally attaining a leading position in freight forwarding.

e. Tastes

Unexplained differences in preferences may finally account for a different composition or sophistication of domestic demand among countries. Some differences, such as the Italian tastes for style and fashion in clothing or the Japanese desire to buy the latest model of consumer electronics, have already been mentioned in Section 3. It is also argued that many Swiss consumers have a tradition of being health conscious asking for natural medicine and herbal products. This in turn may have contributed to a relatively large domestic demand for herbal cough drops. Similarly, some have argued that the Swiss are risk averse and more safety-oriented than others. This would explain why Swiss demand for fire detection and insurance was among the highest in the world giving these industries a boost in exports.

This source of differences in domestic demand can obviously be modelled by assuming different preference systems and thus different utility functions as described by the approaches in Section 2*a*. The dynamic aspects of the sophistication of domestic demand for certain products, i.e., the conspicuous consumption behaviour described above, may be captured by a model assuming interdependent preferences among individuals.

---

[10] As mentioned above, another reason for this lack of success may have been that these firms became important suppliers to the Swiss government (mainly to the military) which indirectly subsidised them.

© Blackwell Publishers Ltd 1996

5. CONCLUSIONS

The presented taxonomy of fundamental demand characteristics and an illustration based on a large set of case studies shows that differences in the composition and sophistication of domestic demand among countries had a major impact on the pattern of international trade in many industries. The discussion of some determinants of domestic demand indicated that differences in demand among countries may not purely arise from different tastes, but include many factors and constraints such as income, relative prices, government regulations or climate and topography.

Most interestingly, dynamic aspects of the mentioned characteristics of domestic demand seem to be especially important. An internationalisation of domestic demand patterns helps domestic firms to achieve a long-term comparative advantage if the industry is characterised by dynamic or external economies of scale. Competition among domestic buyers and learning by using effects seem to establish a fruitful environment for innovation by local supplier firms. Finally, a high growth of domestic demand for a certain product segment is likely to have a positive impact on companies' international competitiveness.

This paper focused on 'domestic' demand, i.e., differences in demand among countries. As is typical for international trade literature, this picture is based on the presumption that countries are and remain distinctive — also with respect to demand characteristics. The examples described above support this view. However, it is also true that the influence from demand is, in many cases, more regional than national in scope. Thus, nearby demand for Swiss firms does not stop at the Swiss border, but may include part of Germany, France and Italy. Similarly, demand characteristics of Western United States or even of the Californian State may be much more important for a firm located in this region than those of national demand. Therefore, modelling demand effects may require to assume transaction costs which are variable with respect to geographic distance and which differ among industries and countries.

Where the differences in demand among countries originate, how they are influenced by government policy, for example, in the area of environment and safety standards or incentives, and how the exact transmission process from domestic demand to domestic production and international trade works, remains to be investigated more carefully. Without a better understanding of this relationship, one can hardly analyse the trade effects which arise from any government policy influencing a country's domestic demand. The role of domestic demand may become even more compelling given that international trade in services is characterised by highly tailored products to suit the individual consumer and expanding.

As shown above, there exist some interesting approaches in international trade theory which could be extended, and many more ways of capturing the described effects from domestic demand may be developed. With respect to empirical research, a more thorough method of linking demand with international trade patterns is required, provided that data availability will allow this.

REFERENCES

Bandelier, D. (1995), *The Swiss Textile Machinery Industry*, in M.J. Enright and R. Weder (eds.).

Bhagwati, J.N. (1982), 'Shifting Comparative Advantage, Protectionist Demands, and Policy Response', in J.N. Bhagwati (ed.), *Import Competition and Response* (Chicago: University of Chicago).

Borner, S., M.E. Porter, R. Weder and M.J. Enright (1991), *Internationale Wettbewerbsvorteile: Ein strategisches Konzept für die Schweiz* (Frankfurt: Campus).

Bumbacher, U. (1995), *The Swiss Watch Industry*, in M.J. Enright and R. Weder (eds.).

Dinopoulos, E. (1988), 'A Formalization of the 'Biological' Model of Trade in Similar Products', *Journal of International Economics*, **25**, 95–110.

Eaton, B.C. and R.G. Lipsey (1989), 'Product Differentiation', in R. Schmalensee and R. Willig (eds.), *Handbook of Industrial Organization, Vol. I* (Amsterdam: North-Holland).

Enright, M.J. (1992), *The Japanese Facsimile Industry in 1990*, Harvard Business School, Case No. 9-391-209.

Enright, M.J. (1995a), *The Swiss Comfort Controls Industry*, in M.J. Enright and R. Weder (eds.).

Enright, M.J. (1995b), *The Swiss Fire Detection Equipment Industry*, in M.J. Enright and R. Weder (eds.).

Enright, M.J. and R. Weder (1995) (eds.), *Studies in Swiss Competitive Advantage* (Bern: Peter Lang).

Falvey, R.E. and H. Kierzkowski (1987), 'Product Quality, Intra-Industry Trade and (Im)perfect Competition', in H. Kierzkowski (ed.), *Protection and Competition in International Trade* (Oxford: Basil Blackwell).

Flam, H. and E. Helpman (1987), 'Vertical Product Differentiation and North-South Trade', *American Economic Review*, **77**, 810–822.

Globerman, S. (1990), 'Canada: Technology and Competitiveness', in H. Soesastro and M. Pangestu (eds.), *Technological Challenge in the Asia-Pacific Economy* (Allen & Unwin: Sydney).

Greenaway, D. (1987), 'The New Theories of Intra-Industry Trade', *Bulletin of Economic Research*, **39**, 95–120.

Grubel, H.G. and P.J. Lloyd (1975), *Intra-industry Trade. The Theory and Measurement of International Trade in Differentiated Products* (New York: John Wiley & Sons).

Hippel, E. von (1988), *The Sources of Innovation* (Oxford: Oxford University Press).

Hunter, L. (1991), 'The Contribution of Nonhomothetic Preferences to Trade', *Journal of International Economics*, **30**, 345–358.

Hunter, L.A. and J.R. Markusen (1988), 'Per Capita Income as a Determinant of Trade', in R.C. Feenstra (ed.), *Empirical Method for International Trade* (Cambridge: MIT Press).

Krugman, P. (1980), 'Scale Economies, Product Differentiation, and the Pattern of Trade', *American Economic Review*, **70**, 950–959.

Linder, S.B. (1961), *An Essay on Trade and Transformation* (New York: John Wiley & Sons).

Markusen, J.R. (1986), 'Explaining the Volume of Trade: An Eclectic Approach', *American Economic Review*, **76**, 1002–1011.

Melvin, J.R. (1993), 'Demand Conditions in International Trade Theory' (University of Waterloo: mimeo).

Porter, M.E. (1990), *The Competitive Advantage of Nations* (London: MacMillan Press).

Richterich, F. (1995), *The Swiss Sugar Confectionery Industry*, in M.J. Enright and R. Weder (eds.).

Rosenberg, N. (1982), *Inside the Black Box. Technology and Economics* (Cambridge: Cambridge

© Blackwell Publishers Ltd 1996

University Press).

Stigler, G.J. and G.S. Becker (1977), 'De Gustibus Non Est Disputandum', *American Economic Review*, **67**, 76–90.

Veblen, T. (1899), *The Theory of the Leisure Class. An Economic Study of Institutions* (London: George Allen & Unwin, Ltd, edition 1924).

Vernon, R. (1966), 'International Investment and International Trade in the Product Cycle', *Quarterly Journal of Economics*, **80**, 190–207.

Weder, R. (1995a), 'Domestic Demand, Comparative Advantage, and the Pattern of Trade', *Schweizerische Zeitschrift für Volkswirtschaft und Statistik* **131**, 377–388.

Weder, R. (1995b), 'Linking Absolute and Comparative Advantage to Intra-Industry Trade Theory', *Review of International Economics*, **3**, 342–354.

Weder, R. (1995c), *The Swiss Dyestuff Industry*, in M.J. Enright and R. Weder (eds.).

12

LIPPINCOTT LIBRARY
DOCUMENT DELIVERY

WHARTON REPROGRAPHICS

International Trade In Services: The Case of India's Computer Software

V.N. Balasubramanyam and Ahalya Balasubramanyam

I. INTRODUCTION

INDIA, with its many paradoxes and puzzles, is never far from the news. For the past four or five years, its fast growing software industry has repeatedly featured in the *Financial Times*, the *Wall Street Journal* and BBC radio and television. The industry has attracted the attention of the media not only because of the fast growth of its exports, at around 40 per cent per annum, but also because a poor country with a per capita income of less than $400 is home to a service industry at the heart of the information technology revolution. Whilst this apparent paradox has its explanations, the presence and growth of India's computer software exports should allay the many fears and doubts of developing countries concerning liberalisation of international trade in services. Indeed, India's computer software industry makes for an excellent case study of the many issues discussed in the literature on services including definitional and conceptual ones, the relevance of trade theory to analysis of trade in services, sources of comparative advantage in services and the welfare implications of trade in services for developing countries.

This paper discusses these and other issues in the context of India's computer software industry. Section 2 sketches the dimensions and structure of the sector in India. Section 3 discusses the service oriented characteristics of software. Section 4 discusses the sources of comparative advantage of India's software industry. Section 5 discusses the welfare implications of the growth of the sector for the Indian economy. Section 6 concludes.

V.N. BALASUBRAMANYAM and AHALYA BALASUBRAMANYAM are from the Department of Economics, Management School, Lancaster University. This paper is part of a research project in Software in South India funded by the British Academy. The authors are grateful to an anonymous referee for helpful comments.

© Blackwell Publishers Ltd 1997, 108 Cowley Road, Oxford OX4 1JF, UK and 350 Main Street, Malden, MA 02148, USA.

WORLD ECONOMY 20(6): 829-844

2. STRUCTURE AND DIMENSIONS

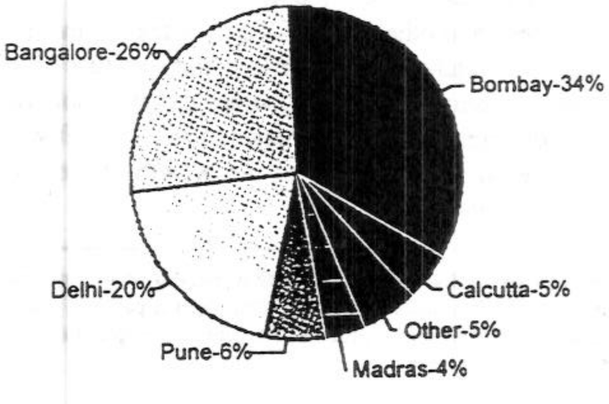
Currently, the total output of India's software industry including in-house development is worth about $750 million, which accounts for only one per cent of the world market. But the sector has registered phenomenal growth at around 40 per cent per annum for the last seven years. The export industry has grown by nearly 47 per cent and the domestic industry by 31 per cent. It is this rapid growth, especially in exports, which has attracted attention.

Estimates of the number of people employed in the sector range from the often cited figure of 1.4 million programmers to 120,000 reported by the National Association of Software and Service Companies (NASSCOM). The former may be an exaggeration while the latter may be an under-estimate. All we can say is that the number of jobs provided by the sector is not insignificant. Nearly 80 per cent of the industry is concentrated in Bombay, Bangalore and Delhi.

The Indian Software Directory (1996) produced by the National Association of Software and Service Companies (NASSCOM) reports that there are about 330 companies in the software export market and other 390 companies which produce software primarily for the domestic market. In addition, there are also hundreds of other start-up companies. Three-quarters of these companies employ fewer than 40 people. At the top end, nearly ten per cent are large companies employing more than 100 technical people each. Around 75 per cent of the companies are either wholly Indian owned or Indian companies with foreign connections, 14 per cent are wholly foreign owned and the rest are joint ventures.

The US accounts for 61 per cent of the export market and European markets for another 17 per cent. Around 70 per cent of the total exports come from the top

FIGURE 1
Location of Software Companies



© Blackwell Publishers Ltd 1997

FIGURE 2
Structure of Ownership

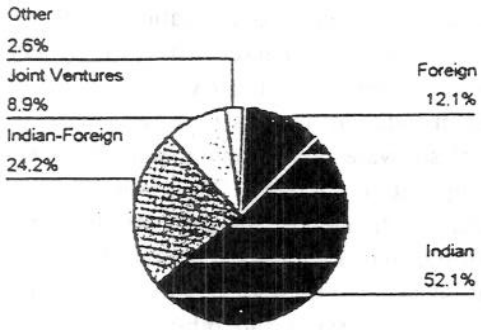
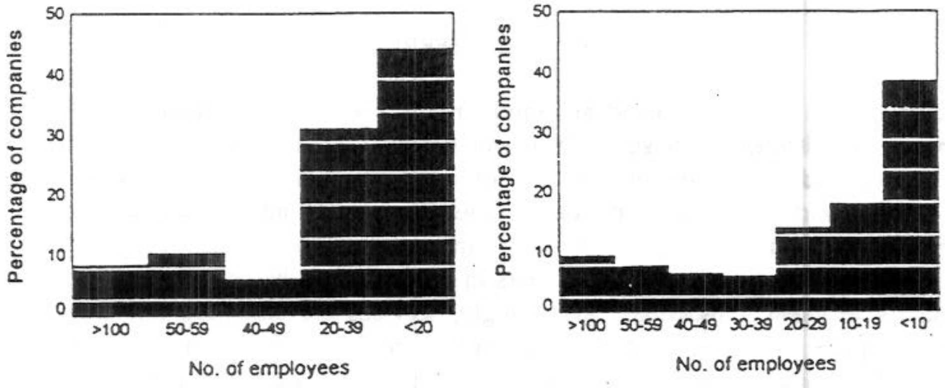


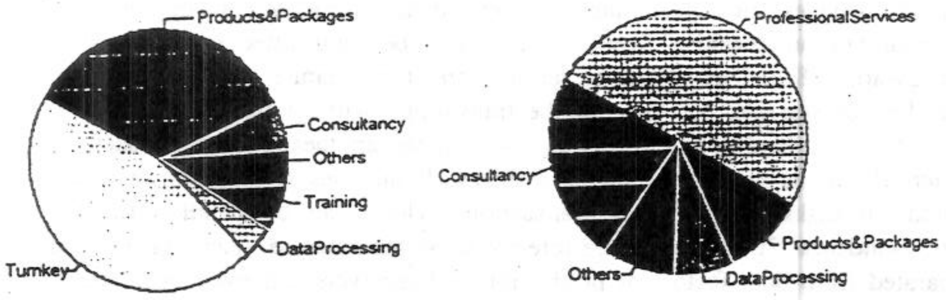
FIGURE 3
Size of Software Companies



Export

Domestic

FIGURE 4
Types of Software Work



© Blackwell Publishers Ltd 1997

20 exporters, which makes for a heavy concentration of the business in the hands of large companies. The share of the market is much more uniformly distributed among the companies in the domestic industry.

Software companies in India are increasingly becoming specialised, focusing on specific segments of software development, with 58 per cent of companies developing end-user application products. The three main sectors which they service now are insurance, financial services including banking and certain segments of manufacturing. Activity, both export and domestic, can be broken up into segments such as turnkey, professional services, products and packages, consultancy, training, data processing and others. About three-quarters of the work done by the export companies is classified as professional services and consultancy. The share of products and packages is ten per cent and data processing eight per cent.

3. IS SOFTWARE A SERVICE?

One of the issues discussed at length in the literature is whether or not services are different from goods. Many of the controversial issues relating to tradability and the mode of delivery of services in international markets revolve around this issue. The often noted distinguishing characteristic of services is their non-storability. It is this intangible nature of services which requires the presence of service providers in the locale of the consumer or vice versa for efficient transactions. Accordingly, services are classified into those which require the presence of the consumer or receiver of the service in the locale of the producer and those which require the presence of the producer in the locale of the receiver. In both cases, efficient transactions require mobility of factors of production and/or mobility of consumers across geographical borders.

There are, however, a number of services which can be delivered without the movement of either the producer or the consumer. These are the so-called 'long distance services' such as live music concerts, data transmission and certain types of insurance and banking services which can be transmitted over the wires (Bhagwati, 1987). These types of services are in the nature of intangibles and may be non-storable, but they can be transmitted without the need for factor mobility — a characteristic of goods. There is yet another class of transactions which also does not require factor mobility. It includes information and data stored on disks or tapes and transactions which can be handled through correspondence. Such services are referred to as separated services, for they are separated from both factors of production and receivers (Sampson and Snape, 1985). As Sampson and Snape observe, in these cases, the demarcation of services from goods is not at all clear.

© Blackwell Publishers Ltd 1997

India's software industry exemplifies these complex definitional issues which have implications for the mode of delivery and policies towards trade in software. Based on the mode of delivery, India's software transactions can be classified into the following categories (Tapasjee Mishra, *Financial Times*, 6 December 1995).

- Software produced overseas: This involves temporary movement of Indian engineers to the location of receivers abroad. Software produced by Indian engineers for the multinational company Reebok at its British headquarters in Lancaster is one example in this category.
- Software produced in India for clients abroad referred to as off-shore production: Indian software engineers resident in India produce software to the specifications of clients abroad or process information supplied by overseas customers. Such off-shore production of software and processed information and data are often transmitted abroad via satellite. In this respect, off-shore production resembles the so called long-distance services.
- Building software products for export overseas: In this case, Indian firms provide an intermediate input for further use by the user abroad.
- Building software products that take advantage of native expertise in an application area such as finance and banking.

These four broad categories of India's software business illustrate the complex nature of service transactions. The third and fourth categories of specialist software can be stored on disks and transmitted abroad much like goods. These types of software can be used as intermediate inputs in a variety of final goods and service industries. The effective utilisation of such software does not require the presence of the producer in the location of the consumer. It is arguable if such software products which are tangible and exported are to be classified as services.

Although India's software companies are increasingly becoming specialised and focusing their production on the last two categories, they do not as yet have a major presence in world markets in these types of software. During the late eighties and the early nineties, much of India's software business was in the first category, but in recent years off-shore production has gained in importance. The switch from production overseas to off-shore production illustrates several features of transactions in services.

First, trade can be a substitute for factor mobility when the latter is subject to restrictions. The switch in mode of delivery seems to be mainly a response to the visa restrictions on temporary movement of labour imposed by receiving countries such as the US. Press reports indicate that:

> Indian engineers find it difficult to obtain the H1B type of visa for temporary residence in the US to begin with, but new restrictions against Indian software specialists' entry into the United States has made the process even more cumbersome (Jain, 1995).

© Blackwell Publishers Ltd 1997

There are also frequent statements by Dewang Mehta, the Executive Director of NASSCOM, concerning visa restrictions imposed by the US. Thus, for Indian software firms, the US market provides 61 per cent of business.[1] But exports of on-site services are now threatened by new US via rules that make it difficult for Indian software engineers to get visas to work on development sites in the United States. The visa restrictions have forced many Indian companies to do the development work at home and then transmit the results via satellite-rather than send their employees overseas (Sharma, 1995).

As stated earlier, transmission of software products and processed information via satellite more nearly resembles the long distance type of service transactions. Indian companies such as *Infosys*, one of the leading software companies located in Bangalore, provide overseas customers with a dedicated software team which, via satellite link, serves as an extension of the operations of the overseas customers using computer facilities which belong to the customers (*Financial Times*, 17 November 1995). Whilst this switch from production overseas to production in India illustrates that restrictions on factor movements may not be a total barrier to trade in services, it is arguable if such long distance delivery of services is as efficient as production in the locale of the receiver. Two-way communication between the supplier and receiver of services via satellite may serve to reduce inefficiencies in delivery, but this feature may be specific to trade in software.

Second, with the growth in restrictions on movement of labour, capital appears to have moved to the location of labour. As stated earlier, there is a significant presence of foreign owned companies in India's' software industry. Examples of such firms are Hewlett-Packard, IBM, Texas Instruments and Motorola, all of which have production facilities in Bangalore. Whilst US immigration laws are a factor in the movement of capital to labour, the newly liberalised trade and investment regime of the government of India and various sorts of incentives provided by the government for the sector, including duty free imports of hardware, have also encouraged foreign investment in the sector.

Third, in the case of services such as software, movement of capital to labour may, in fact, serve to preserve efficiency of service transactions if not to enhance it. Foreign firms in the sector not only provide capital, but they also perform management and marketing functions which serve to promote efficiency. Also, in many cases, the software produced by foreign firms may be for their own use, especially so in the case of overseas production which is relocated in India because of restrictions on labour mobility. In these cases, it is the customer or receiver of services who moves to the location of the producer and the product is produced to the specifications of the consumer, a sort of bespoke software.

---

[1] It is reported that in the mid-nineties, 72 per cent of software exports from India was in the form of professional services in which firms send software engineers to work at clients' sites overseas.

© Blackwell Publishers Ltd 1997

India's software sector illustrates that services can be delivered in many ways and as in the case of goods, trade can often be a substitute for factor mobility. Software, however, displays one of the distinguishing characteristics of services in general noted by Hill (1977). It is that:

> the production of a service cannot generally be distinguished from that of a good by means of the technology used, but by the fact that the producer unit operates directly on goods which already belong to the consumer of the service.

Both in the case of data processing and bespoke software referred to earlier, the producer operates directly on the data and information provided by the consumer. An example of such software is the processing of accounts, including tickets and booking, of airlines such as Swissair and British Airways.

4. WHY SOFTWARE IN INDIA?

It is a puzzle that a developing country with a per capita income of less than $500 and an adult literacy rate of only 50 per cent should have a recognisable presence in international markets for a human-capital intensive service product such as software. It is, of course, well known that although it is an abysmally poor country, India possess a large pool of highly trained engineers and scientists. It also has the second largest higher education system in the World. There are 4 million students enrolled in 7,000 colleges and 150 Universities (Altbach, 1993). It is tempting to conclude that India possesses a comparative advantage in the human-capital intensive software industry because of her endowments of human capital. But why software and not the many other human-capital intensive manufacturing industries? Also India is not the only developing country with endowments of human capital, South Korea and the Philippines too, are endowed with an educated pool of labour, but they do not possess as visible a presence as India in the international markets for software.

There are several economic and sociological explanations for the birth and rapid growth of software in India. First, the endowments of human capital referred to above, a product of India's education policies over the years which emphasised investments in tertiary education. Second, is the contribution of expatriate Indians to the growth of the sector. Third, several managers of software firms attribute India's comparative advantage in software to the innate mathematical abilities of Indians.

Admittedly, the availability of a large pool of highly educated young Indians is a factor in the growth of the sector. young Indian engineers trained in the prestigious Indian Institutes of Technology seek work in the sector because of the relatively high salaries it offers, the attractive working conditions, training facilities and the opportunities for travel which it provides. Typically, salaries for

© Blackwell Publishers Ltd 1997

software engineers in their early twenties is around 50 per cent higher than those in academia and the manufacturing sector. Moreover, in most of the software companies, the sort of hierarchical management structures typical of traditional manufacturing industries in India seem to be conspicuous by their absence. Many of the companies in Bangalore have successfully imported new and innovative work practices such as the formation of learning cells and group work designed to promote problem solving through collaboration and co-operation.

The pool of trained engineers India possesses provides a substantial cost advantage for the sector. Importers of software in the US and the UK report cost savings of around 40 to 50 per cent compared with what they would have to pay for comparable services at home. As stated earlier, such cost savings have influenced companies such as North West Water in the UK and international airlines such as Swissair and British Airways to outsource their software to locations in India. It should be added that the quality of software produced by Indian companies is highly rated and apart from cost savings, it is the ability of Indian engineers to deliver high quality software efficiently which is a major factor in the growth of the industry. In this context the frequently heard remark that Indians possess an innate aptitude for mathematics, after all it was India which contributed the numeral zero (shunya in Sanskrit) to the world of Mathematics, is worth recording.

Whilst India can claim a comparative advantage in the production of software on various grounds, the government has assisted the sector in various ways to exploit the advantage. The liberalisation of trade and industrial licensing policies in 1991 resulted in the simplification of procedures for setting up new businesses and shortened start-up time for both domestic and foreign firms. Policies of specific advantage to the sector include the setting up of Export Processing Zones, Software Technology Parks which can import hardware duty free, exemptions from sales and excise taxes and subsidies for rent, power and water.

Although much of the evidence is impressionistic, yester years brain drain appears to have been a factor of importance in the establishment and growth of the sector. Indian engineers and technicians who migrated to the west, mostly to the US, during the sixties and the seventies appear to be a source of finance, technical assistance and market intelligence to the Indian companies. Several of the 40 companies in Bangalore we interviewed were established by return migrants. Examples include Wipro Limited, Infomart and BPL Systems. In the case of others, expatriate Indians had provided the finance and market intelligence. Many of the return migrants who had worked in information technology industries in the US appear to have transplanted management techniques they had learned abroad. There are also firms which are established by Indians who reside abroad but visit India frequently — the so called 'to and fro migrants'. The sector has also facilitated the reverse brain drain by providing not only jobs, but also an investment outlet for the savings accumulated by the expatriates.

© Blackwell Publishers Ltd 1997

All of these factors have been instrumental in providing the requisite finance, expertise and market information for the software sector. Even so, it can be asked 'why was the software sector the focus of the expatriate Indians?' They could have targeted many other segments of the manufacturing and service sectors of the Indian economy.

The structural characteristics of the sector provide an explanation. First, it does not require extensive overheads, save a link to the export markets which is provided by satellite facilities installed by the government. Second, most expatriates who have returned to India are familiar with the work habits and work culture abroad, which facilitates the so called body shop software end of the business. Third, and most important, the sector enables Indians to keep abreast of the developments in the industry because of the ease with which information can be transmitted via the satellite link. Indeed, the nature of the business is such that the engineers have to communicate with their clients almost continuously. This they can do from the confines of their offices. Finally the sector provides a unique opportunity for engineers and technicians to exploit a comparative advantage in conceptual work.

In sum, the sector is almost made to order for trained Indians who are now used to the work ethic and work culture of the west. Indian expatriates and Indian engineers who are trained abroad are able to move in both the Indian and western cultures with ease. Also, the nature of the industry is such that entrepreneurs are not dependent on unreliable suppliers of components and parts for they need virtually none. They are not compelled to suffer delays and red tape characteristic of most business activities in India, including licences for imported components and payment of bills. They mostly cater to the export markets which promptly pay their bills. And the government of India has eased their access to imported hardware through its liberalised trade policy.

5. IMPACT OF THE INDUSTRY

Despite its rapid growth, India's software sector is still in its infancy — it accounts for only one per cent of the world market for software. As such, any judgements on the welfare implications of the sector for the economy may be premature. The available information on the extent, nature and structure of the sector, however, permits a set of hypotheses or guarded speculations. A frequently heard remark in India is that the software sector is the leading sector. It is seen as a force of liberation from the low level equilibrium trap in which poor countries such as India are trapped. At the other extreme is the view that the software sector is no more than a small island of affluence and growth in a vast sea of poverty and backwardness — a view which was expressed to us by both businessmen and academics more than once in Trivandrum in the State of Kerala.

© Blackwell Publishers Ltd 1997

Both views represent the extremes — one is born out of euphoria and a sense of national pride and the other out of cynicism and the traditional hostility to private enterprise in the state of Kerala. If there is a middle ground, of what does it consist?

India can be justly proud of its software sector, which belongs to the high tech group and has attracted not only much attention, but is also seen as a threat by leading players in the US. The visa restrictions on Indian software engineers imposed by the US is a sign that the US perceives the Indian industry as a threat to the US industry. Indeed, the software sector seems to be a public consumption good of the sort identified by Johnson (1965), in the sense that apart from providing private benefits such as employment, it promotes national pride in much the same fashion as victory in a Cricket Test match or the Olympics. The fact that the BBC should have launched more than one programme on the sector and the *Financial Times* should have devoted a special section to the sector, feeds national pride in India. And to the extent that public consumption promotes social welfare, the computer software sector has made a major contribution to the Indian economy.

Apart from its role as a public consumption good, the sector does contribute to net foreign exchange earnings. It is primarily an export industry which is not dependent on imported components in a major way except for imports of hardware. Such imports are in the nature of fixed costs rather than variable costs.

In addition, the sector provides employment to the educated and trained Indians, which in itself is a major contribution. Indeed, firms at the less technologically intensive end of the spectrum of the industry provide employment to a substantial number of young Indians with a first degree. And many of these employees are women. The wage rate for employees in these firms exceeds their opportunity costs by a substantial margin. And the work environment which is clean and free of pollution of any sort is an additional bonus. The firms in this group are mainly engaged in data processing and data entry.

Another feature of the sector which should be counted as a gain to the economy is the formation and promotion of entrepreneurship and innovation. The young entrepreneurs in the industry are a new breed whose style of management differs markedly from that in the traditional industries. The communication channels between the top management and the young software engineers appears to be free of the hierarchy and bureaucracy typical of the traditional industries. The managers are acutely conscious of the need for training and the promotion of job satisfaction for their employees. It is likely that the structure and characteristics of the sector compels enlightened management. The workforce consists of relatively young people — the average age of software engineers is around 25 years — they are educated and open to new ideas. The export markets are highly competitive and do not allow for delays in delivery or shoddy

© Blackwell Publishers Ltd 1997

workmanship and the turnover of labour in most firms, around 20–25 per cent, is relatively high. Such a high turnover compels managers to institute attractive incentive schemes designed to retain labour. Profit sharing schemes, worker participation in both profits and management, opportunities for short spells of work abroad are all common features in most firms. The manager of one of the companies in Bangalore told us that he was thinking of turning the brain drain problem to his advantage. He was thinking of dangling the carrot of US green card status, which he would obtain, after the engineers had served his firm for a stipulated period! In addition, the training and life styles which the managers have acquired abroad, many of whom are part of the reverse brain drain referred to earlier, also explains the enlightened style of management. In sum, the sector makes a diverse range of contributions to the economy.

The other side of the case, however, should also be put. Here, there are two sets of arguments. First, the sector has generated an internal brain drain. Second, the sector has resulted in a phenomenon which can be termed 'brain drain without emigration'. A sub-set of this argument is that the sector is no more than an export enclave.

The internal brain drain thesis is that the sector is creaming off the best brains in India. Brilliant young engineers, physicists and mathematicians are lured to the sector because of the high wage rates, good working environment and the life style it offers. But are these young professionals overqualified for the sort of jobs they do? Does the writing of software really require highly trained engineers from the Indian Institutes of Technology? Some academics we spoke to expressed the opinion that this group of people constitute a loss to the academic and research world. Also, the sector had led to a misallocation of educational resources to the extent that a host of institutions offering degrees and diplomas in computer sciences and electronics are springing up in the country. In sum, the rest of the economy is deprived of human capital because of the existence of the sector which competes effectively for trained labour.

These propositions cannot be empirically tested in the absence of detailed information on the specific educational requirements of the sector and the nature of the human capital it is attracting. There are also other factors which complicate the issue. One of these is that the software sector is not homogenous producing an undifferentiated product. It includes at one extreme, elementary tasks such as data entry, which may not strictly speaking constitute software business, to the technically sophisticated task of producing software packages. In between there are a range of activities including data processing and consultancy. In addition there are companies which are dedicated to the task of producing software for specific sectors, such as banking and finance. In essence, the sector consists of a group of non-competing firms, each of which has its own niche market. The requirements of labour of each of these firms may differ considerably. To say that all of them are creaming off the best brains in the country may be an exaggeration.

© Blackwell Publishers Ltd 1997

In strict economic terms, the judgement on the issue has to be based on the social opportunity costs of trained labour in India. Here again, labour in the sector cannot be treated as one homogenous group. For instance, the opportunity cost of labour, mostly female labour, in data entry may be close to zero. In the absence of employment in the sector, labour engaged in these activities may be unemployed. In this case, the charge of internal brain drain may be fallacious. Admittedly, many of those performing repetitive data entry jobs, such as the ones in the firm in Kerala which edits the Yellow pages of British Telecom, possess first degrees and may be somewhat overqualified for the jobs they do. They are hired in preference to those who do not possess first degrees, but can adequately do the job, because of excess supply of labour in the market. Faced with an excess supply and a fixed minimum wage they have to pay, employers prefer those with a degree to those without a degree. This is because of the notion that it is fair to employ those who have invested in education as opposed to those who have not undertaken such investments (Bhagwati and Srinivasan, 1977). Here, it is the excess supply of labour and the sticky wage which results in the hiring of over qualified people, who are more easily trained and whose private rates of return to education may be higher than the social rate of return, and the blame cannot be laid at the door of the software sector.

The thesis of the internal brain drain though is addressed to jobs at the top end of the spectrum in the software sector. The young engineers in this segment of the sector are paid salaries in excess of their opportunity cost. An academic teaching Physics or Mathematics, in one of the Universities earns much less than a software professional of a similar age. Computer specialists who are graduates of prestigious institutions, working in the top companies, earn as much as $1000/- per month. This is a considerable sum and has an element of rent. Do wage rates in academia and the manufacturing sector represent true opportunity costs of work in the software sector? Software is a highly competitive business and calls for skills and a work ethic which is demanding. Also, there seems to be an excess demand for software engineers. It is estimated that in 1997, the total shortage in the number of qualified people will have grown to almost 50,000 in the domestic and the export markets. It appears that the education system is yet to respond fully to the demands placed by the sector on it. In the absence of a sufficient number of young graduates trained specifically for the industry, managers of software firms are forced to employ graduates with higher degrees in science and engineering. The competition for these graduates is not so much from the domestic manufacturing sector, but from job markets abroad. The brain drain seems to impose high costs on the sector. Indeed, there are those who argue that the sector has contributed to the brain drain by providing on the job training and information about job markets abroad. There may be a germ of truth in the internal brain drain thesis, but it is again imperfections in the labour market and an education system which is yet to cater to the needs of the sector that are to blame.

© Blackwell Publishers Ltd 1997

The argument that the sector may be no more than an export enclave and that it appears to have promoted brain drain without emigration has echoes of Singer's (1950) thesis, enunciated in the early fifties, concerning foreign investment in developing countries. We have discussed this thesis in the context of India's software sector in detail elsewhere (Balasubramanyam and Balasubramanyam, 1996) and the discussion here is confined to the salient points. It is a fact that by its very nature, the sector forges very little backward and forward linkages with the rest of the economy. It does draw labour from the local economy and its employees do spend their incomes on goods and services in the economy. But the sector, by the very nature of its business, requires very little by way of locally produced inputs. Although there is no data to estimate the extent of the sector's backward and forward linkages, it is safe to say that its secondary employment and income effects tend to be relatively low. Again, the expenditure patterns of the young software engineers who earn relatively high incomes are likely to be concentrated on the luxury goods and services end of the domestic market. Such expenditures are unlikely to be import intensive, in any case, the government continues to impose steep quotas and tariffs on imports of consumer durables. Whether or not such expenditures generate strong secondary employment effects is open to question. It is also a fact that in Bangalore, the presence of the sector has escalated real estate prices; such inflation of house prices has no doubt served to widen income disparities in the city.

Young software engineers not only earn above average incomes and generally enjoy a life style which is distinctly upper class, but they also have opportunities to travel abroad. Moreover, because of the nature of their work, they are in constant contact with their counterparts abroad via satellite. Also, they are a distinct professional group who commune with each other and revel in the 'computer speak' of bytes and roms, frequently move from one company to another in search of status and higher salaries with the ultimate destination of most of them being the US. It is arguable if all this suggests that brain drain has not ended, it has taken a new form — 'brain drain without emigration.'

Finally, it is to be noted that software is an input and not a final product. It contributes to growth of productivity and efficiency of the activities in which it is used. The sector could be of assistance to Indian producers of final goods and services in coping with increased competition in the post-liberalisation era. As yet, India's manufacturing sector is unspecialised, much too diverse and for the most part, characterised by low growth in productive efficiency, all of which are consequences of the import-substituting industrialisation strategy the country pursued until recently. New technology in the form of indigenously produced software could enhance productive efficiency and help transform India's manufacturing sector.

© Blackwell Publishers Ltd 1997

Admittedly, computerisation raises the spectre of unemployment. Effective utilisation of software which would promote efficiency requires investment in education and training, in the absence of which implementation of computerisation may be counterproductive. Once India develops a viable domestic market, the sector could make a valuable contribution to the economy and its export segment will be on a firm footing. If India were to export software packages, it has to promote a broadly based sophisticated domestic market. There are a number of theories which suggest that the export market is often an extension of the domestic market. But, to repeat, the onus of developing a domestic market rests not only on the companies, but also on the education establishments in the country.

6. CONCLUSIONS

India's software sector, because of its rapid growth, has attracted much discussion and the attention of the media both in India and abroad. The presence of a technologically progressive service industry in a developing country is seen by some as somewhat of a puzzle. This paper has analysed the reasons for the presence and growth of the software sector. It has also reviewed its likely economic impact on the Indian economy. Three of the conclusions of the paper are noteworthy. First, the reasons for the presence of the software sector are to be sought not only in economics but also in sociology. A proximate reason for the growth of the sector in India has to do with the pattern of India's education investments and the brain drain India experienced during the sixties and the seventies. Indian expatriates abroad appear to have provided the initial stimulus for its growth. Second, the paper suggests that if the sector is to experience sustained growth based on exports of sophisticated products, it has to be grounded in a viable domestic market for its products. Growth of the domestic market, however, requires considerable investments in education and on the job training in the use of computers in India's manufacturing sector and service producers such as banking and insurance. It is worth remembering that mechanisation may displace unskilled labour, but it generates jobs for skilled labour. Third, the paper examines the arguments that the software sector may be no more than an export enclave and that it may also have generates an internal brain drain. There may be some truth in these arguments, but here again the problem seems to lie with the Indian education system. It seems to be lopsided in the sense that it produces graduates who cannot find jobs for the qualifications they possess. The education job ladder adjusts to the oversupply of educated persons by employing people in jobs for which they may be overqualified.

REFERENCES

Altbach, P.G. (1993), 'The Dilemma of Change in Indian Higher Education', *Higher Education*, 23, 1.

Balasubramanyam, V.N. and A. Balasubramanyam (1997), 'The Distribution of Gains Between the Investing and Borrowing Countries Revisited, the Case of the Emerging Indian Software Sector', in D. Sapsford and John-ren-Chen (ed.), *Economic Development and Policy* (Macmillan, forthcoming).

Bhagwati, J. (1987), 'Trade in Services and the Multilateral Trade Negotiations', *The World Bank Economic Review*, 1, 4 (September).

Bhagwati, J.N. and T.N. Srinivasan (1977), 'Education in "Job Ladder" Model and the Fairness-in-Hiring Rule', *Journal of Public Economics*, 7, 1–22.

Jain, A. (1995), 'News Analysis: The Software Industry', *News India*.

Johnson, H.G. (1965), 'An Economic Theory of Protectionism, Tariff Bargaining, and the Formation of Customs Unions', *Journal of Political Economy*, LXXIII, 3, 327–45.

Sampson, G.P. and R.H. Snape (1985), 'Identifying the Issues in Trade in Services', *The World Economy*, 8, 171–82.

Sharma, R. (1995), 'Economic Watch: Software Industry Growing', *News India*.

Singer, H.W. (1950), 'The Distribution of Gains Between Investing and Borrowing Countries', *American Economic Review, Papers and Proceedings* (May).

© Blackwell Publishers Ltd 1997

WHARTON REPROGRAPHICS

Environment and Planning A, 1993, volume 25, pages 1103–1114

'It's gotta be da shoes': domestic manufacturing, international subcontracting, and the production of athletic footwear

R Barff
Department of Geography, Dartmouth College, 6017 Fairchild Hall, Hanover, NH 03755-3571, USA
J Austen
34 Bradford Road, Wellesley, MA 02181, USA
Received 2 January 1993

Abstract. An apparent paradox exists in the geography of athletic footwear production. The industry is highly labor intensive; it is also locationally volatile. Although most production now takes place in Southeast Asia, the United States nevertheless still contains significant enclaves of athletic footwear production. Moreover, the cheapest shoes continue to be produced in the United States whereas more complex, expensive models tend to be manufactured in Asia. To understand this geography, we must move beyond the basic consideration of international labor-cost differentials. By means of two case studies, it is shown that domestic production involves very different labor processes from those of production based in other countries and, like many other sectors of the economy, domestic producers gain advantage by carrying smaller inventories through faster lead times. The best explanation, though, centers on the shoes themselves. Athletic shoes produced in the United States tend to have many fewer stitches in them than those manufactured elsewhere, which minimizes the most expensive component of the production process. Furthermore, tariffs on athletic shoes massively discriminate against imported shoes of a particular construction.

1 Introduction

1.1 *Overview*

The athletic footwear industry, by any measure, is highly dynamic. In economic terms, both sales and volume of production grew very rapidly in the 1980s. Sales of athletic footwear in the United States probably tripled during the last ten years (Willigan, 1992) and, according to one study, Reebok International Ltd has been the fastest growing company in the history of US business (Korzeniewicz, 1992). The industry is also volatile geographically. For example, in 1989 Nike Inc. (the brand with the largest market share in the United States) had about 2% of their shoes made by Chinese-based subcontractors: in 1993, almost a quarter of their shoes will come from Chinese factories. More generally, in the last thirty years, the locus of production within South and East Asia has shifted radically (Donaghu and Barff, 1990; Levy, 1991). At a different spatial scale, many producers of athletic footwear have developed a complex set of long-term and short-term subcontracting agreements with other firms that change from year to year as a result of factory improvements, market fluctuations, and technological progress (Donaghu and Barff, 1990).

Although source countries continue to change, the majority of athletic footwear production continues to occur in the Southeast Asian region. The three US companies which account for over 60% of sales in the USA have the vast majority of their production there. Several athletic footwear firms, however, still manufacture in the United States. This type of athletic footwear company generates particular interest because domestic production persists despite huge discrepancies in labor costs between the USA and abroad. Based largely on two detailed case studies, this research explains the dichotomy in the geography of athletic footwear production within the context of contemporary theory on the spatial division of labor. We argue that athletic footwear production represents an industry that cannot be exclusively characterized by analysis of comparative labor costs and a spatial division of labor at an international scale, but rather is best understood by a deeper theorization of processes operating at various geographic scales—from the firm to the region, nation, and eventually to supranational entities. Specifically, this involves unearthing the relations between labor costs, the nature of the manufacturing process, market strategy, and regulatory policy.

The emphasis on regulatory policy, particularly tariffs, separates this paper from much previous Anglo research on international trade dynamics and the new international division of labor (NIDL), which has, in the main, neglected the effects of tariffs on industrial location (see McConnell, 1986; O'Loughlin, 1990; Rees, 1992). By way of contrast, regulatory policy in the form of tariffs, quotas, and exchange controls in Latin America (for example, Gwynne 1986), the Caribbean (Schoepfle and Perez-Lopez, 1989), and in other parts of the Third World (Gereffi and Wyman, 1990) are almost always considered as part of an assessment of the course of industrial development. Given the interest in geographical restructuring in the United States, regional adjustment, and regulatory frameworks (for example, Clark, 1989; 1990; 1991; 1993), we seek to broaden that focus to include international regulatory policy. This effort may be especially timely given that, at the time of writing—summer 1992, a North American Free Trade Agreement (NAFTA) had just been signed between Canada, the United States, and Mexico and the Multi Fiber Arrangement (MFA) in textiles and apparel expires. Increasingly, economic geographers will have to make more explicit consideration of trade and tariff policies (Rees, 1992).

1.2 *Restructuring in the shoe industry and the new international division of labor*

Shoe manufacturing (like some other nondurables) was once a major part of the economies of the world capitalist core. In the United States, employment in the shoe industry has declined steadily during the 20th century as mass production in vertically integrated facilities gradually replaced the vertically disintegrated craft production of the previous centuries. To a certain extent, production also shifted geographically to the South and Midwest, but Massachusetts in particular and New England in general remain the focus of the industry today. Since about 1970, the US footwear industry has struggled even more as low-skilled standardized production decentralized to the global periphery. In the last decade alone, employment in nonrubber footwear production has dropped from 143000 to about 90000 in 1990 while imports of shoes have more than doubled (Footwear Industries of America, 1989). This sector of the US economy is particularly vulnerable to competition from nations with low average manufacturing wages (Donaghu and Barff, 1990; Parkinson, 1990).

One of the simplest and most powerful explanations for this dispersal based on wage rates centers the well-known NIDL theory (Fröbel et al, 1980; Hymer, 1972; Massey, 1984; Schoenberger, 1988). Originally conceived as a treatment of the role of the multinational corporation in uneven development, the NIDL idea put forward the view that multinationals are endowed with internal divisions of labor, commodity flows, and hierarchical modes of organization (Walker, 1989). Whereas the old horizontal division of labor placed the periphery in a role supplying primary goods and raw materials for processing in the global core, NIDL theory emphasized a different geography where new pressures forced corporations to restructure and seek cheaper sites of production in the periphery. Improvements in transportation and communication, the evolution of a large supply of disposable labor in developing nations, and the increasing subdivision of modern labor processes facilitated this dispersal, and was partially caused by heightened competition between producers in traditional core countries and those in newly industrializing countries in the periphery.

Despite (and perhaps because of) its simplicity, the NIDL thesis has become one of the main tools in explaining recent trends in global industrialization (for example, Dicken, 1992). In this paper, of course, we look specifically at footwear production: an archetype of the NIDL model. In many parts of the world, the footwear industry, like the clothing industry, continues to be highly labor intensive, reliant on narrow divisions of labor, and most shoes sold in the United States are manufactured in developing countries. The NIDL model, however, fails to explain the tenacious presence of producers in the United States where labor costs far exceed those in many developing countries.

2 Methodology

At the beginning of this project, we set out to investigate in detail three or four US-based athletic footwear companies. Of the several we contacted, two (Converse Inc. and New Balance Athletic Shoe Inc.) allowed access to the factory and senior management and workers for interview. The interviews, carried out in the summer and autumn of 1990, provided much of the empirical information for this research. The unstandardized format of the interviews yielded special insight into the locational and production strategies of the companies in the context of a variety of contingent circumstances (McDowell, 1992; Schoenberger, 1991). Where necessary, quantitative data from published sources supplement the qualitative information we derived from interview. Blending both qualitative and quantitative data allowed us to develop an empirically rich contextual analysis of US-based producers of athletic footwear.

3 The athletic footwear industry

3.1 *Market share*

A clear hierarchy in the athletic footwear industry has evolved in the United States. Nike and Reebok hold the largest market share and dominate sales, and constitute the first tier in the hierarchy. They subcontract most of their production to factories in Southeast Asia and also market the widest range of athletic footwear. Second-tier firms, which, on the basis of 1991 market share data of branded shoes, include LA Gear Inc., Converse, Keds Corporation, and Adidas AG, possess a smaller market share than the first-tier firms (figure 1). Nevertheless, these companies

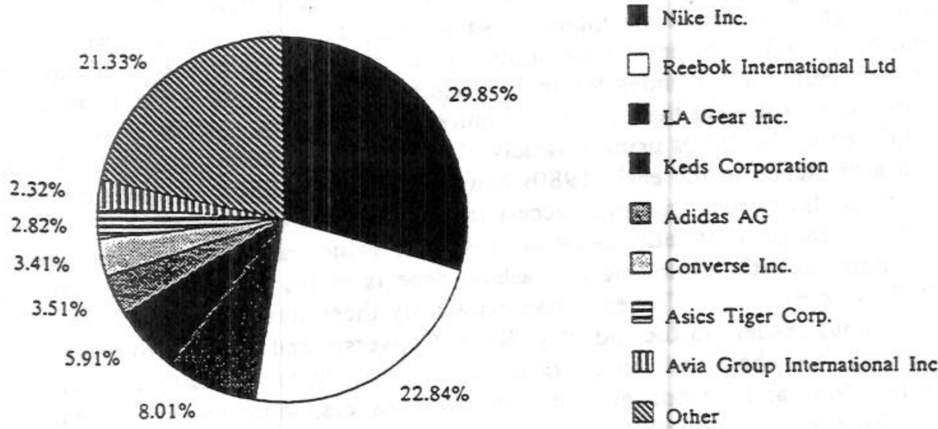


Figure 1. 1991 Market share of branded shoe sales in the United States (source: Sporting Goods Inc., 1992).

produce a substantial annual volume of shoes in a variety of market areas. Third-tier firms—the other 18% of the US market—generate significantly smaller volumes of shoes in a narrower range, and may be characterized as niche producers, specializing in a particular type of athletic shoe. Nike serves as an exemplar of the first tier producers and Donaghu and Barff's (1990) study shed some light on the nature of its production system. One goal of this study is to complement the work of Donaghu and Barff and illuminate the production strategies and geography of smaller producers in what we are calling the second and third tiers. Converse serves to exemplify second-tier production; New Balance was chosen as illustrative of a third-tier producer.

The introduction of the McKay sewing machine in 1852 caused a revolution in production in the footwear industry in general (Hoover, 1937; Melder, 1979). Converse, a rubber company, began mass production of machine-stitched athletic shoes in 1908. Within about a decade, the company was producing the shoe it continues to be famous for—the All-Star. Without any notable change in design, over the next fifty years the All-Star became the premier athletic shoe; a status obtained, for example, by being selected as the official athletic shoe of the US Olympic basketball team in 1936 and being chosen as the official athletic shoe of twenty different US Olympic teams in 1972.

Converse's dominance of athletic footwear production eroded rapidly in the 1960s and 1970s. New players like Nike, Vans Inc., Asics Tiger Corp., Hi-Tec Sports, USA, Etonic Inc., and others emerged, marketing performance athletic footwear with new types of cushion and support mechanisms. During the 1970s, Americans began to pay more attention to health and fitness and millions began to participate in sports such as jogging and aerobics. As a result, athletic footwear producers designed and marketed new sneakers for the purpose of enhancing performance while reducing the risk of injury: the athletic footwear industry reinvented itself as one catering to a new leisure society. Consequent to the fitness boom, the general popularity of athletic footwear grew as Americans increasingly found the comfort of athletic shoes appealing and their look fashionable (Donaghu and Barff, 1990; Rikert and Christensen, 1984).

Since this point of 'takeoff' about twenty years ago, the athletic footwear industry changed dramatically. As the market expanded, not only did the number of athletic footwear companies increase, but also the number of models they produced. Today, new product lines are introduced seasonally, including 'back to school' shoes and products issued during the National Collegiate Athletic Association basketball tournament in the spring. Firms have developed athletic shoes for every purpose from weight lifting to street hiking to windsurfing to cross training. Despite this trend of specialization in athletic footwear, no more than about 20% of consumers use their performance shoes for their designated purpose. As a result, marketing has become critical in the industry (Ramirez, 1990). Nike rose to a position of dominance in the 1970s using a variety of successful marketing strategies. Reebok's rapid ascendence in the early 1980s coincided with the boom in aerobic fitness: Reebok parlayed their massive success in that segment of the market into success across the range of athletic footwear. LA Gear gained significant market presence even more recently, exploiting the fashion aspects of the industry more than Nike and Reebok ever did. The explosive growth by these three companies relegated the previous leaders in the industry (Keds, Converse, and Adidas) to a much smaller market share with an attendant smaller role in the industry as a whole.

While Nike and Reebok grew into industry leaders, some smaller companies have also gained a foothold in the athletic shoe industry. For example, New Balance focuses exclusively on particular niches of the athletic footwear market—

producing only running, walking, and tennis shoes, with running shoes accounting for about 55% of total sales. Moreover, New Balance holds a special niche in its specific markets because it provides the consumer with a considerable choice of size: for example, it remains the only athletic footwear firm that produces shoes that vary in width from AA to EEEE, and uses this to advertise itself.

3.2 *Production geography*

The notable shifts in market share in the athletic footwear industry were accompanied by equally dramatic locational shifts in production. Nike, Reebok, and LA Gear subcontract the lion's share of their production to manufacturers in Southeast Asia and this part of the world presently serves as the principal locus of production for the US market. Furthermore, the geography of production within Southeast Asia is hardly static. In the early 1970s, Japan served as the main center of production and is crucial to understanding the present-day international production geography.(1) For the last fifteen years, manufacturers in South Korea and Taiwan have dominated the US market. Presently, the center of gravity of production is now shifting markedly toward China, Thailand, and Indonesia and, if this trend continues, these countries will be the main export platforms for Nike and Reebok in the next decade.

In contrast to the first-tier firms, many second-tier and third-tier firms, especially those that are US owned, continue to produce significant numbers of shoes at US facilities. Some interesting geographic shifts in production have taken place both at the regional and at the national scales, however. Converse began production in Malden, MA, a small city immediately north of Boston and in the heart of the then robust US footwear industry. As the All-Star's reputation grew, Converse gradually expanded production to four other sites in New England. This pattern of development mirrored that of many nonathletic footwear manufacturers (Estall, 1966; Hoover, 1937).

The company continues to behave like many other textile and footwear producers. In the last twenty years, Converse closed the New England factories and opened three new factories: one in Lumberton, NC, one in Mexico, and the other in Puerto Rico. In 1989 the company shut down the Puerto Rican plant and consolidated eastern US production at the plant in North Carolina. New Balance's factories are, and have been since its founding in 1906, in the Boston area and southern Maine and situated, for the most part, in inner cities (Pereira, 1992). Although these second-tier and third tier producers continue to manufacture shoes in the United States, they are not immune to the pressures to exploit the cheap labor locations in newly industrializing countries and the Third World. Both Converse and New Balance market shoes made in Southeast Asia. They both also use foreign sources to manufacture components for final assembly in the United States. Nevertheless, both companies presently pride themselves on being US companies and the bulk of their shoes are mass produced in the United States. Furthermore, the Converse All-Star is one of the simplest types of athletic shoes and yet rather than subcontracting the production of this shoe offshore, Converse issues subcontracts only for the much more complex and expensive leather and nylon shoes it markets.

4 The persistence of US athletic footwear production

4.1 *Labor*

The consolidation of Converse's US production facilities in Lumberton in the early 1970s coincided with the emergence of stiff new competition from producers manufacturing shoes in Southeast Asia and parts of the Third World. Whereas companies such as Nike looked for inexpensive nonunionized labor locations in the global periphery, Converse sought a similar place in the periphery of the global core. North Carolina as a whole had the lowest average hourly earnings in manufacturing in the United States in the 1980s (US Bureau of Labor Statistics, 1989). Additionally, rural Lumberton is in the poorest county in North Carolina and the majority of the nonunionized 1200 workers in the Converse factory are native American women. Converse's factory thus relies on a work force in a captive labor market many would hold up as emblematic of the global 'periphery'. That said, the wages paid in Lumberton at the Converse factory are still about thirty times greater than the pay at the footwear factory in Tangerang, Indonesia, which presently manufacturers top-of-the-line Nike Air shoes.

---

(1) Countries in Central and South America had a tradition of casual shoe and boot production and continue to do so. Asian countries have developed a tradition for both athletic and casual footwear manufacturing.

Despite such wage differentials between the United States and Southeast Asia, Converse remains competitive in this environment in some obvious ways. Workers in their US plant tend to be more productive because they are trained for six months and tend to stay with the factory much longer than workers in Southeast Asia. In contrast, Asian laborers—especially in Thailand, Malaysia, and China—work long hours under poor working conditions and consequently have high rates of turnover (Frank, 1981; Jenkins, 1984; Landsberg, 1979). Productivity gains have also been achieved by the process of line-balancing—the grouping of particular workers together on the basis of speed in order to maximize efficiency and minimize the lag time between specific tasks performed by workers. With this system of production, a piece-rate method of pay, typical of the footwear industry in general, is not possible. Enhanced training periods and the elimination of a piece-rate pay scales both reduce labor turnover. Workers are less likely to be discouraged in the first few weeks and months of employment when their relative skills tend to be inferior to those others who have worked longer.

New Balance also has a six-month training period for its workers. Minorities make up about 50% of the work force at New Balance, a large proportion of whom are new immigrants. Lessons are in English, Spanish, and Portuguese and New Balance claims their retention rates are very high.

Apart from relatively extensive training, New Balance uses other means to increase efficiency. The company uses modular team manufacturing (a type of flexible labor practice), wherein workers are trained to perform many different functions and can adapt to accomplish necessary tasks. Workers are arranged into teams each of which is responsible for the production of a specific model of shoe. All the tasks required in shoe assembly occur in one place on the factory floor and each team member can perform lasting, roughing, sole laying, and packing. Interteam competition is encouraged by the company. New Balance calculates that this method of organization has reduced the production time of an average size shipment of shoes from longer than a week to less than a day. This system has not only improved efficiency, it also has resulted in fewer defects and fewer returns from retailers and customers. This system of production is particularly effective for New Balance given that the company offers shoes in more widths than any other. This variety requires shorter production runs—something more easily accommodated in the relatively small factories New Balance operates than in the factories reliant on scale economies and Fordist techniques used by first-tier producers in Southeast Asia.

4.2 *Shoe construction*

Despite differences in productivity, factor supply considerations remain a key component in manufacturing athletic footwear and domestic producers attempt to minimize labor costs in several ways. Stitchers receive some of the highest wages in the factory. Labor constitutes approximately 40% of the production costs of each shoe, of which stitchers make up the majority of the cost. In other words, if stitching is minimized, then labor costs are reduced. One way to minimize stitching cost is to produce less performance-oriented, simpler athletic shoes that require fewer stitches. Such sneakers tend to be canvas shoes with basic designs that do not include padding, ankle supports, or decorative features that necessitate many stitches. Converse's All-Star and Jack Purcell shoes typify this type of athletic footwear.

The construction of the All-Star consists of a simple, minimally reinforced canvas upper lasted to a rubber sole. There is no excess stitching on the upper as even the Converse emblem is laminated onto the canvas rather than sewn. The stitching operations consist only of the staying of the uppers and side pieces, and the binding of the quarters and tongues. (That said, even this 'minimal' stitching represents seventeen to nineteen separate operations per shoe.) Keds and Tretorn AB, two other producers of fashion-oriented canvas shoes, also manufacture some their product in the United States. Vans manufacture their canvas shoes exclusively in Vista, CA.

Domestic producers have worked to design shoes that reduce labor costs in production, yet US-based manufacturers employ other tactics to remain cost competitive. Although Converse and New Balance maintain some stitchers at their domestic production facilities, (over 100 machines in the case of Converse), they conduct much of their stitching outside US borders. Converse stitches nearly all the uppers for its canvas shoes at a facility in Reynosa, Mexico. The wages paid at the Mexican facility are about one fifth those offered at the Lumberton plant. The Reynosa factory is also a maquiladora, so Converse pays no tariffs on the shoe uppers produced there. Being in northeastern Mexico, Reynosa offers a location from which Converse can transport stitched and unstitched upper back and forth relatively easily and still take advantage of significant savings in labor costs (South, 1990).

In contrast, New Balance subcontracts the stitching of some uppers to firms in South Korea. This feature of their production strategy fits well with the assertion that high-performance shoes require more stitching and thus cost more to produce, as the uppers stitched offshore tend to be parts of New Balance's most sophisticated models. The subcontracting arrangement benefits New Balance by allowing them to take advantage of relatively inexpensive Asian labor, but also in limiting import tariffs. Tariffs are lower on parts, like uppers, than on fully assembled shoes. New Balance also subcontracts stitching to a firm in Saco, ME, which suggests that specialization subcontracting even within the USA reduces production costs significantly.

The fact that New Balance makes shoes of varying widths adds another dimension to this discussion of shoe construction and competitive position. Such flexibility is not economically feasible in many Asian plants that subcontract with first-tier producers. There, factories run on tighter schedules with one set of cutting dies, one set of lasts, and one set of outsole molds for each model. Production takes place in fast, long runs where workers cannot make adjustment for widths.

4.3 *Service requirements*

The athletic footwear industry, like that of footwear in general and other sectors including clothing and automobiles, now pays much more attention to lead times, delivery speed, and inventory control than ever before. As market competition and consumer demands increase, manufacturers are forced to provide more services to the retailer to maintain price stability. The nature of the service requirements frequently varies by the particular market within an industry. In the athletic footwear industry, the service requirements for high-performance athletic shoes can be met equally well by firms manufacturing in the USA or abroad, but the service requirements for more fashion-oriented and low-performance athletic shoes can be better fulfilled in the United States. This distinction occurs in the way athletic shoes are designed and marketed. The high-performance shoe, whether or not it is used for its given purpose, is designed and sold on the basis of comfort, the attractiveness of specialty features, and the styling. Furthermore, in the United States, high-performance shoes are frequently marketed by celebrities and often sell on the basis of whose name endorses them. Ironically, the expense of high-performance shoes may also lure the customer, as expensive athletic shoes are perceived by some as status symbols. In contrast, lesser-performance athletic shoes sell on the basis of their low price, and simple aesthetic features such as color. Because these shoes are inexpensive, simple in design, and not associated with celebrities or a new technology, consumer preferences tend to change relatively rapidly.

As a result of market shifts, the impact of service requirements on the low cost segment of the athletic shoe market is turnaround time in production. As in other sectors of the economy, the retailer and manufacturers monitor the colors and styles selling the best during a particular period, and on short notice, respond to meet the demand of these trends (compare Parkinson, 1990). This need for the rapid production and distribution of a particular shoe favors manufacturing in the United States rather than subcontracting to Asia. In contrast, the market for high-cost athletic shoes does not necessitate service requirements that favor any particular manufacturing location. Rather, the requirements focus on promoting a particular shoe through endorsements and making the retailer aware of the special attributes of a shoe to be able to sell more of them. Nike and Reebok, in selecting high-profile celebrities to endorse their shoes and through consistent design innovation, essentially define trends in high-performance athletic footwear as opposed to responding to them.

Converse and New Balance serve as examples of this distinction between the high-cost and low-cost market segments. For instance, Converse has become better able to minimize inventory and accelerate turnaround times in production by rating the quality of its retailers and by using an improved communication network between the company and its retailers. The closing of the facility in Canovanas, Puerto Rico, epitomizes Converse's effort to speed production turnaround. Despite the loss of a source that employed relatively inexpensive labor, closing this plant enhanced manufacturing and shipping efficiency through the consolidation of production and the elimination of split orders for single customers. Domestic producers can meet immediate retailer demand effectively with low inventory and a type of just-in-time production, whereas companies subcontracting offshore meet new demands for a particular style or color of shoe through an intricate system of flexible subcontracting (Donaghu and Barff, 1990) and/or having larger inventories: that is, tending more to just-in-case production (Markides and Berg, 1988).

New Balance, as a manufacturer of relatively high-cost, high performance athletic shoes, must promote the technical features of their product and educate retailers in how best to explain these features to the customer. Yet offshore producers such as Nike and Reebok undertake sales and marketing programs equally as well. In fact, because first-tier producers have considerably more capital than privately held smaller companies like New Balance, they employ more people in marketing and sales and create far more extensive programs and promotions. New Balance is left with one option: they concentrate on a select group of retailers, with whom they establish close relations. These retailers become knowledgeable about New Balance shoes and promote the company.

4.4 *Real regulation*

National-level regulatory policies can define the international geography of an industry and can be as powerful an influence on industrial location as is labor (Lipietz, 1986).[2] For instance, following the example in many technologically advanced nations, development policies in Latin America (and elsewhere) have relied on protectionist measures to shield local producers from foreign competition (Gwynne, 1986, chapter 3). One prevailing wisdom behind such behavior was that tariffs and import restrictions helped nurture infant local industry and only after a country developed a mature industrial structure would a country relax these restrictions and trade more freely.

Although the major free trade agreements presently being negotiated in Europe and North America help focus attention on trade liberalization, technologically advanced nations, like many in the 'South,' still maintain an intricate set of restrictions that apply to a wide range of industries. In the United States, tariff restrictions (contained in the country's 'Harmonized Tariff Schedule' or HTS) undergo continuous revision. For athletic footwear, the HTS currently determines to a large extent what types of shoes are profitably manufactured in the country and what types are best produced offshore. Essentially, two factors are at work. The first is the cost of a shoe or subassembled shoe component; the less expensive the shoe or component, the greater the impact of the tariff on import cost in terms of percentage increase per shoe. This occurs because tariffs frequently consist of a fixed dollar amount plus a percentage. For some components, the dollar amount alone might exceed the value of the product—see below. The second is the grading system of tariffs based on materials, construction, and final use; the greater the tariff on a particular shoe, the less profitable it becomes to manufacture that shoe offshore. The HTS is extremely complicated, occupying two large volumes which detail almost 9000 different rates. Within the footwear section of the HTS, the grading system remains complex, but for our purposes a general discussion suffices.

One of the most important issues in tariff variability is the manner in which the upper is attached to the sole of the shoe. There are three main types: nonmolded, molded, and fox-banded (where a strip of material is attached around where the sole and the upper meet in order to reinforce the bond). Nonmolded shoes carry the lowest duty, fox-banded the highest. The materials of the upper are also critical to the assessment of the tariff, which increases from synthetic leather, through genuine leather, nylon and leather combined, to textiles. Variations in sole attachment combined with variations in upper material can produce large variation in tariff from the 6% applied to men's synthetic leather nonmolded shoes, to the 8.5% on men's leather basketball shoes, to the 10.5% on nylon mesh running shoes to the 37.5%-plus-90 cents on canvas fox-banded shoes.

These differences in tariffs affect the geography of athletic footwear production. In the case of Converse, the geographic organization of production appears to be a direct result of the tariff system. Converse manufactures all its fox-banded canvas shoe models in the United States and subcontracts the production of its leather molded basketball shoes abroad. Furthermore, the cost of the shoe augments the duties on these two types of footwear: canvas athletic shoes are much cheaper to produce than leather ones, so a tariff represents a greater percentage of the total cost of a canvas shoe than of a leather shoe. Note also that the canvas uppers for Converse were stitched in a tariff-free maquiladora. Tariffs for imported textile uppers range from 9%-plus-30 cents to 11.2%-plus-30 cents. These tariffs are relatively large considering the importer's cost per stitched upper is about 25 cents.

---

(2) For an account of tariff and nontariff protection of the Australian footwear, clothing, and textile industries, see Capling and Galligan (1991).

Tariffs influence the geography of New Balance's production too. The company produces its running and court shoes (footwear with mesh, leather, and nylon uppers) in Maine and Massachusetts. Leather basketball and aerobic shoe production is subcontracted to firms in South Korea and Taiwan. Tariffs on basketball shoes average 8.5%; tariffs on running shoes start at 10.5%. Recall also that New Balance import stitched upper for several models of running shoe. They do this because the tariff on complete running shoes is over 10%: no tariffs are applied to imported nylon or leather uppers alone. New Balance therefore uses subcontractors to bypass the most expensive part of athletic footwear production (stitching) without having to pay tariffs.

5 Conclusions

Offshore manufacturing occurs primarily in those industries where technological or economic considerations inhibit mechanization (Jenkins, 1984). In the athletic footwear industry, production methods have not changed significantly in decades and the major producers subcontract work to manufacturing firms in Southeast Asia. Nevertheless, Converse and New Balance, like some other companies, produce domestically, yet they take advantage of the NIDL by having one phase of their production process situated offshore instead of their complete operations. At the international scale, then, this challenges Hansen's assertion (1979) that manufacturers involved in labor-intensive industries maintain cost competitiveness by shifting their geographic location of production. At a different spatial scale, however, Converse and New Balance behave as Hansen predicts. Converse and New Balance both incorporate disenfranchised labor pools within the United States, a tendency that parallels the strategy of many other companies, in a range of industries, to relocate outside of the United States to take advantage of much lower labor rates in other parts of the world.

The fact that the geographic production strategies of Converse and New Balance adhere so closely to the structure of the US tariff system attests to the power of regulatory policy in shaping the global geography of athletic footwear production. US production of athletic footwear might not take place without the HTS. The details of the system of tariff allocation has built up over decades and the 8753 different rates in the HTS have been described as "equivalent to over 8000 different industrial policies and 8000 different entitlement programs for protected domestic industries" (Bovard, 1991, page 8).(3) The impact of regulatory policy for athletic footwear provides support for the assertion that the analysis of industrial location and the spatial division of labor must move beyond factor prices and cost minimization (Schoenberger, 1988). As protectionist pressures increase with adverse macro-economic conditions in an industry (Takacs, 1981), more research should explicitly account for the potential impact of tariffs and other protectionist measures on the location of industry at the global scale (compare Aw and Roberts, 1986; Chang, 1987; Winters and Brenton, 1991).

Although companies such as Converse and New Balance appear set to produce athletic footwear domestically for some time, the potential ratification of an NAFTA hardly assures the future of companies like Converse and New Balance. While the final form of the agreement has yet to be established, it is hard to speculate about its specific impact. Suffice it to say, as Mexico has established producers and a tradition of the production of nonathletic footwear, Mexican-based production at present is probably a much greater threat to the dress and casual footwear industries than to US-based athletic footwear manufacturers. Furthermore, given the morphology of subcontracting between first-tier companies like Reebok and Nike and their production partners in Asia, the future of Mexican production is at least partially dependent on the spatial investment strategies of South Korean and Taiwanese investors (Donaghu and Barff, 1990). The length of time Converse and New Balance will continue to produce in the USA if NAFTA is ratified relates also to the negotiated phaseout period of tariffs and the specification of the rules of origin from Mexico. Other technologically advanced countries have enacted elaborate labor-adjustment schemes and industrial redevelopment policies in tandem with tariff and nontariff barrier reduction (compare Capling and Galligan, 1991). NAFTA pays lip service to retraining and labor displacement, but has no regionally based, comprehensive scheme to assist specific industries affected by the trade agreement, so barring any major shift in domestic policy, it may be a simple matter of time before US-based production of some types of athletic footwear becomes even more precarious. For a company like Converse, whose system of production established seems built in part around the present structure of import regulations and which has refocused production radically twice the last two decades, the elimination of the regulatory framework may make a new Mexican location irresistible.

---

(3) This system is even more complicated when we take into account nontariff barriers such 'voluntary export restrictions' and similar export-restraint arrangements.

Acknowledgements. This research was supported in part by a grant from the Richter Memorial Trust. Adrian Bailey, Vicky Lawson, and Tom Klak provided some very helpful commentary on previous drafts.

References

Aw B Y, Roberts M J, 1986, "Measuring quality change in quota-constrained import markets" *Journal of International Economics* **21** 45–60

Bovard J, 1991 *The Fair Trade Fraud* (St Martin's Press, New York)

Capling M A, Galligan B, 1991, "Sectoral corporatism with a difference: the textiles, clothing and footwear plan" *Australian Journal of Public Policy* **50** 48–60

Chang E T, 1987, "Protective effects of discriminatory quantitative restrictions" *International Economic Journal* **1** 15–28

Clark G L, 1989, "Remaking the map of corporate capitalism: the arbitrage economy of the 1990s" *Environment and Planning A* **21** 997–1000

Clark G L, 1990, "Restructuring, workers' pension rights and the law" *Environment and Planning A* **22** 149–168

Clark G L, 1991, "Limits of statutory responses to corporate restructuring illustrated with reference to plant closing legislation" *Economic Geography* **67** 22–41

Clark G L, 1993, "Costs and prices, corporate competitive strategies, and regions" *Environment and Planning A* **25** 5–26

Dicken P, 1992 *Global Shift: The Internationalization of Economic Activity* 2nd edition (Guilford Press, New York)

Donaghu M T, Barff R, 1990, "Nike just did it: international subcontracting, flexibility and athletic footwear production" *Regional Studies* **24** 537–552

Estall R C, 1966 *New England: A Study in Industrial Adjustment* (Praeger, New York)

Footwear Industries of America, 1989 *Footwear Manual* Footwear Industries of America, 1420K Street NW, Suite 600, Washington, DC 20005

Frank A G, 1981 *Crisis in the Third World* (Heinemann Educational Books, London)

Fröbel F, Heinrichs J, Kreye O, 1980 *The New International Division of Labor* (Cambridge University Press, Cambridge)

Gereffi G, Wyman D L, 1990 *Manufacturing Miracles: Paths of Industrialization in Latin America and East Asia* (Princeton University Press, Princeton, NJ)

Gwynne R N, 1986 *Industrialization and Urbanization in Latin America* (The Johns Hopkins University Press, Baltimore, MD)

Hansen N, 1979, "The new international division of labor and manufacturing decentralization in the United States" *Review of Regional Studies* **9** 1–11

Hoover E M, 1937 *Location Theory and the Shoe and Leather Industries* (Harvard University Press, Cambridge, MA)

Hymer S, 1972, "The multinational corporation and the law of uneven development", in *Economics and World Order* Ed. J Bhagwati (The Free Press, New York) pp 113 - 140

Jenkins R, 1984, "Divisions over the international division of labor" *Capital and Class* **22** 28 - 57

Korzeniewicz M, 1992, "Commodity chains and marketing strategies: Nike and the global athletic footwear industry", paper presented at the XVI meeting of the Political Economy of the World System section of the American Sociological Association, Duke University, April; copy available from the author of Department of Sociology, University of New Mexico, Albuquerque, NM

Landsberg M, 1979, "Export-led industrialization in the Third World: manufacturing imperialism" *The Review of Radical Political Economics* **11**(4) 50 - 63

Levy B, 1991, "Transactions costs, the size of firms, and industrial policy" *Journal of Development Economics* **34** 151 - 178

Lipietz A, 1986, "New tendencies in the new internatioinal division of labor: regimes of accumulation and modes of regulation", in *Production, Work, Territory* Eds A Scott, M Storper (Allen and Unwin, Winchester, MA) pp 16 - 40

McConnell J E, 1986, "Geography of internatioinal trade" *Progress in Human Geography* **10** 471 - 483

McDowell L, 1992, "Valid games? A response to Erica Schoenberger" *The Professional Geographer* **44** 212 - 215

Markides C C, Berg N, 1988, "Manufacturing offshore is bad business" *Harvard Business Review* **66** 113 - 120

Massey D, 1984 *Spatial Divisions of Labor* (Methuen, New York)

Melder K, 1979, "Life and times in Shoe City: the shoe workers of Lynn", Essex Institute, Peabody and Essex Museum, East India Square, Salem, MA 71970

O'Loughlin J, 1990, "Political geography: attempting to understand a changing world order" *Progress in Human Geography* **14** 420 - 433

Parkinson C, 1990, "Restructuring in the textile, clothing and footwear industries" *Australian Journal of Public Administration* **49** 359 - 361

Pereira J, 1992, "New Balance disproves myth about urban workers" *Wall Street Journal* 9 June, page B4

Ramirez R, 1990, "The pedestrian sneaker makes a come back" *New York Times* 14 October, page F17

Rees J, 1992, "Regional development and policy under turbulence" *Progress in Human Geography* **16** 223 - 231

Rikert D, Christensen R, 1984, "There is no finish line: a Nike case study", unpublished manuscript, Harvard Business School, Cambridge, MA

Schoenberger E, 1988, "Multinational corporations and the new international division of labor: a critical appraisal" *International Regioinal Science Review* **11** 105 - 119

Schoenberger E, 1991, "The corporate interview as a research method in economic geography" *The Professional Geographer* **43** 180 - 189

Schoepfle G K, Perez-Lopez J F, 1989, "Export assembly operations in Mexico and the Caribbean" *Journal of Interamerican Studies and World Affairs* **31** 131 - 161

South R B, 1990, "Transnational maquiladora location" *Annals of the Association of American Geographers* **80** 549 - 570

Sporting Goods Inc., 1992 *Sporting Goods Intelligence Newsletter* Sporting Goods Inc., 23 East Washington Street, Newcastle, PA 16101

Takacs W, 1981, "Pressures for protectionism: an empirical analysis" *Economic Inquiry* **19** 687 - 693

US Bureau of Labor Statistics, 1989 *Employment and Earnings* (US Government Printing Office, Washington, DC)

Walker R, 1989, "A requiem for corporate geography" *Geografiska Annaler B* **71** 43 - 68

Willigan G E, 1992, "High performance marketing: an interview with Nike's Phil Knight" *Harvard Business Review* **70** 91 - 101

Winters L A, Brenton P A, 1991, "Quantifying the economic effects of non-tariff barriers: the case of UK footwear" *Kyklos* **44** 71 - 92

© 1993 a Pion publication printed in Great Britain

BEST ORIGINAL AVAILABLE

14

WHARTON/REPROGRAPHICS

Agile brokers are making fortunes off the spreads between luxury-car prices around the world.

Carbitrage

By Nina Munk

AT THE DOWNTOWN Jeep/Eagle dealership in Toronto a Grand Cherokee Limited V-8 lists for the equivalent of US$28,125. Across the border, in the Buffalo suburb of West Seneca, the same model with identical options carries a sticker price of $32,590. Across the ocean, in Munich, Germany, the price is $56,000.

On these spreads dealers with international connections are making a bundle. Typical is Wolfgang Trenz. Based in Weiherhammer, about 112 miles north of Munich, Trenz's American Motors Trenz GmbH sells the Grand Cherokee Limited for about $46,342, 17% less than the official price in Germany. Stocking almost exclusively Jeeps and Voyager minivans, Trenz last year sold 1,000 new cars and 200 used cars. Trenz says he made about $3 million, on sales of $36 million.

How can he undercut the other dealers? Trenz is a carbitrager. Bypassing Chrysler distribution channels, he buys Jeeps wherever they are cheapest, paying retail but knowing he can sell them for much more. These days the cheapest place is Canada, where the deutsche mark has climbed by 58% against the Canadian dollar in the past four years.

By telephone and fax, Trenz taps into his network of Chrysler dealers in Ontario, Quebec and Nova Scotia. Lately he's paid a Canadian dealer sticker, or $28,125, for a Grand Cherokee Limited. Transportation and Germany's 10% duty and 15% sales tax add $10,000 to the cost. Trenz installs new lights, adjusts the electrical system and changes the tires to meet German vehicle specifications and overrides a computer chip that limits the North American Cherokee's top speed to 112 miles an hour. Finally, Trenz sets aside $1,100 per vehicle to cover warranty costs.

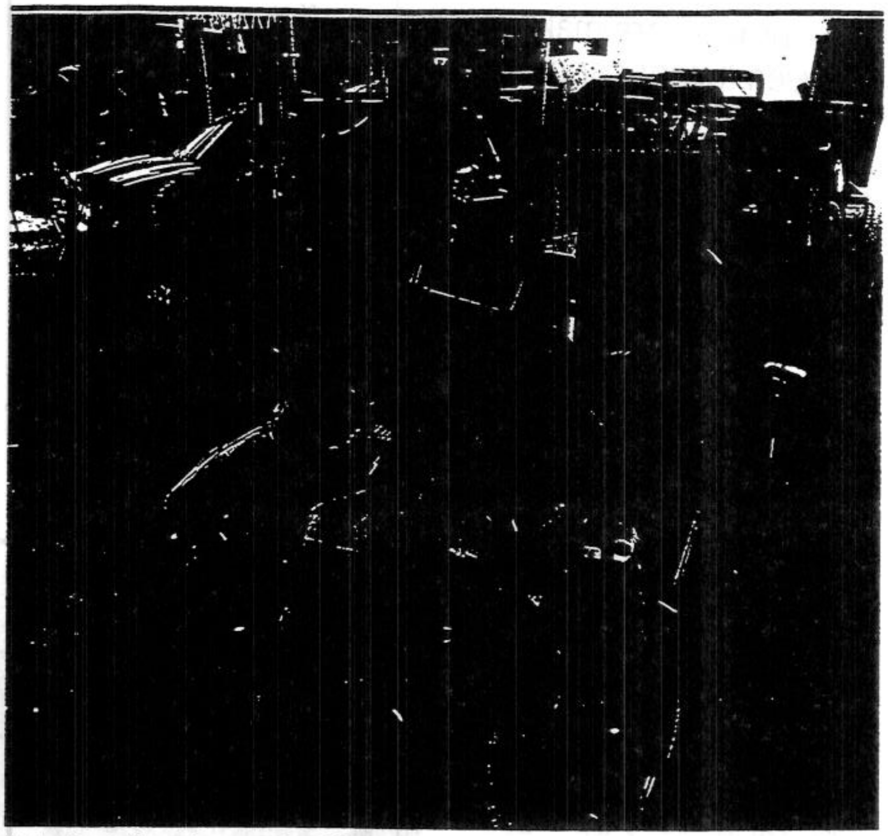
All this brings the total cost of Trenz's Cherokee to $42,972, leaving a profit of $3,370. Trenz says he moves 60% of his metal to Germans, the balance to buyers in Russia, Poland, the Czech Republic, South Africa, Turkey, Zaire and Kazakhstan. Sales, he says, have doubled in each of the past four years.

There are plenty of Trenzes playing intercountry car price spreads. Magazines like Germany's *Auto Motor und Sport* are packed with advertisements

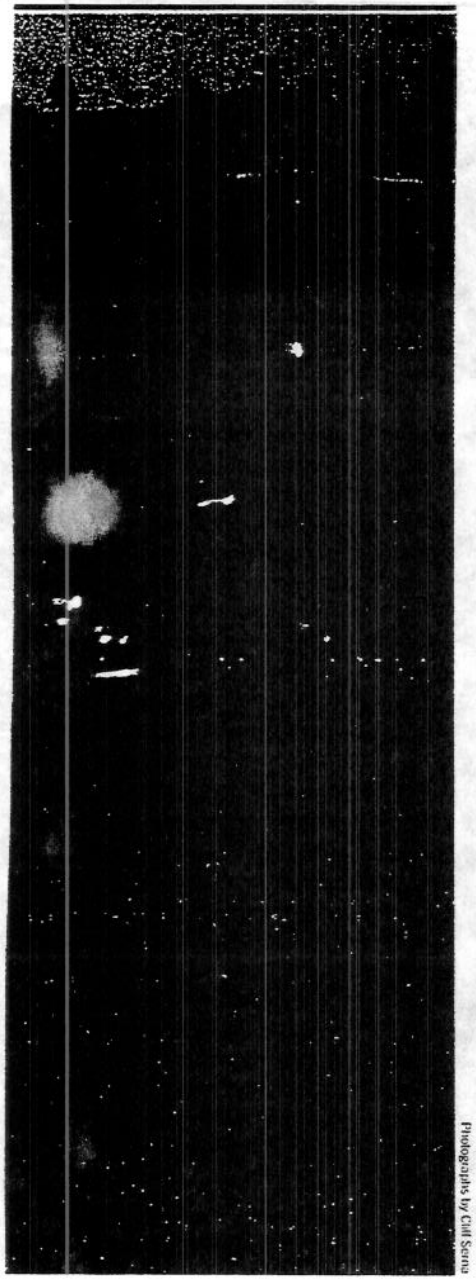
offering American utility vehicles and minivans at big discounts.

And not just American cars. In Russia a BMW 525i purchased through an official BMW dealer runs about $73,755, versus $38,800—before taxes—in Germany. Wolfgang Vollath, managing director of BMW's Moscow office, says Russia's gray market BMW dealers run splashy television ads and outsell his licensed dealers 10-to-1.

Taiwan is a hot market for gray market Volvos, Mercedes and BMWs. Thanks to tariffs and taxes on imported luxury cars, a Volvo 850 Turbo Wagon that retails at $33,550 in Los Angeles goes for $62,535 in Taipei. One gray market broker crisscrosses Canada and the U.S. looking for the models he needs while his Taiwanese partner drums up customers. Within ten days of an order coming in, the cars are delivered to the closest port, ready for shipment to Taipei.



"I don't only have to compete with Chrysler dealers. I have to be something better," says Wolfgang Trenz. Above, Trenz mechanics alter Jeeps to meet German specs.



Photographs by Cliff Serna

"We've got this business down to a science," the broker says. "I'd buy ten times as many Mercedes if I could find them."

The auto manufacturers are not amused. The gray market cars don't hurt their sales, but they do hurt their franchised dealers—and perhaps their image. "These [gray market] vehicles are not made for the markets they're going to," grumbles Philip Bates, manager for international sales analysis at Chrysler. "When there's a problem, they end up at the [official] dealer, but there's no warranty, sometimes no parts, and no way to contact them for a recall. It's a nightmare."

However, the gray marketers have developed aftersales services of their own. About 18 months ago Wolfgang Trenz introduced his own guarantee program, which duplicates Chrysler's official warranty. His service department in Weiherhammer carries so many Chrysler spare parts (purchased on the gray market) that official Chrysler dealers often come to Trenz for parts.

Manufacturers are trying to crack down on the gray market trade. Last fall GM instituted a policy asking that Canadian dealers make all purchasers sign a document promising not to resell their new car outside the country. In July Volvo of North America announced it will fine U.S. and Canadian Volvo dealers $2,000 for each car a dealer sells to an exporter. Canadian dealers report that Chrysler, Ford and GM threaten to cancel franchises of dealers known to be trading with the carbitragers.

But arbitrage is a hard thing to stop. "It's tough to resist a big sale, especially when you've got a lot of inventory on the ground," confides a Canadian Chrysler dealer who has sold to the gray market. He adds: "And unless you're blatant about it, it's pretty hard to get caught." ■

LIPPINCOTT LIBRARY
DOCUMENT DELIVERY SERVICE

INTERNATIONAL INVESTMENT AND INTERNATIONAL TRADE IN THE PRODUCT CYCLE *

RAYMOND VERNON

Location of new products, 191. — The maturing product, 196. — The standardized product, 202.

Anyone who has sought to understand the shifts in international trade and international investment over the past twenty years has chafed from time to time under an acute sense of the inadequacy of the available analytical tools. While the comparative cost concept and other basic concepts have rarely failed to provide some help, they have usually carried the analyst only a very little way toward adequate understanding. For the most part, it has been necessary to formulate new concepts in order to explore issues such as the strengths and limitations of import substitution in the development process, the implications of common market arrangements for trade and investment, the underlying reasons for the Leontief paradox, and other critical issues of the day.

As theorists have groped for some more efficient tools, there has been a flowering in international trade and capital theory. But the very proliferation of theory has increased the urgency of the search for unifying concepts. It is doubtful that we shall find many propositions that can match the simplicity, power, and universality of application of the theory of comparative advantage and the international equilibrating mechanism; but unless the search for better tools goes on, the usefulness of economic theory for the solution of problems in international trade and capital movements will probably decline.

The present paper deals with one promising line of generalization and synthesis which seems to me to have been somewhat neglected by the main stream of trade theory. It puts less emphasis upon comparative cost doctrine and more upon the timing of innovation, the effects of scale economies, and the roles of ignorance and uncertainty in influencing trade patterns. It is an approach with respectable sponsorship, deriving bits and pieces of its inspiration from the writings of such persons as Williams, Kindleberger, MacDougall, Hoffmeyer, and Burenstam-Linder.[1]

Emphases of this sort seem first to have appeared when economists were searching for an explanation of what looked like a persistent, structural shortage of dollars in the world. When the shortage proved ephemeral in the late 1950's, many of the ideas which the shortage had stimulated were tossed overboard as prima facie wrong.[2] Nevertheless, one cannot be exposed to the main currents of international trade for very long without feeling that any theory which neglected the roles of innovation, scale, ignorance and uncertainty would be incomplete.

LOCATION OF NEW PRODUCTS

We begin with the assumption that the enterprises in any one of the advanced countries of the world are not distinguishably different from those in any other advanced country, in terms of their access to scientific knowledge and their capacity to comprehend scientific principles.[3] All of them, we may safely assume, can secure access to the knowledge that exists in the physical, chemical and biological sciences. These sciences at times may be difficult, but they are rarely occult.

It is a mistake to assume, however, that equal access to scientific principles in all the advanced countries means equal probability of the application of these principles in the generation of new products. There is ordinarily a large gap between the knowledge of a scientific principle and the embodiment of the principle in

---

* The preparation of this article was financed in part by a grant from the Ford Foundation to the Harvard Business School to support a study of the implications of United States foreign direct investment. This paper is a by-product of the hypothesis-building stage of the study.

[1] J. H. Williams, "The Theory of International Trade Reconsidered," reprinted as Chap. 2 in his *Postwar Monetary Plans and Other Essays* (Oxford: Basil Blackwell, 1947); C. P. Kindleberger, *The Dollar Shortage* (New York: Wiley, 1950); Erik Hoffmeyer, *Dollar Shortage* (Amsterdam: North-Holland, 1958); Sir Donald MacDougall, *The World Dollar Problem* (London: Macmillan, 1957); Staffan Burenstam-Linder, *An Essay on Trade and Transformation* (Uppsala: Almqvist & Wicksells, 1961).

[2] The best summary of the state of trade theory that has come to my attention in recent years is J. Bhagwati, "The Pure Theory of International Trade," *Economic Journal*, LXXIV (Mar. 1964), 1–84. Bhagwati refers obliquely to some of the theories which concern us here; but they receive much less attention than I think they deserve.

[3] Some of the account that follows will be found in greatly truncated form in my "The Trade Expansion Act in Perspective," in *Emerging Concepts in Marketing*, Proceedings of the American Marketing Association, December 1962, pp. 384–89. The elaboration here owes a good deal to the perceptive work of Se'ev Hirsch, summarized in his unpublished doctoral thesis, "Location of Industry and International Competitiveness," Harvard Business School, 1965.

WHARTON REPROGRAPHICS

a marketable product. An entrepreneur usually has to intervene to accept the risks involved in testing whether the gap can be bridged.

If all entrepreneurs, wherever located, could be presumed to be equally conscious of and equally responsive to all entrepreneurial opportunities, wherever they arose, the classical view of the dominant role of price in resource allocation might be highly relevant. There is good reason to believe, however, that the entrepreneur's consciousness of and responsiveness to opportunity are a function of ease of communication; and further, that ease of communication is a function of geographical proximity.[4] Accordingly, we abandon the powerful simplifying notion that knowledge is a universal free good, and introduce it as an independent variable in the decision to trade or to invest.

The fact that the search for knowledge is an inseparable part of the decision-making process and that relative ease of access to knowledge can profoundly affect the outcome are now reasonably well established through empirical research.[5] One implication of that fact is that producers in any market are more likely to be aware of the possibility of introducing new products in that market than producers located elsewhere would be.

The United States market offers certain unique kinds of opportunities to those who are in a position to be aware of them.

First, the United States market consists of consumers with an average income which is higher (except for a few anomalies like Kuwait) than that in any other national market — twice as high as that of Western Europe, for instance. Wherever there was a chance to offer a new product responsive to wants at high levels of income, this chance would presumably first be apparent to someone in a position to observe the United States market.

Second, the United States market is characterized by high unit labor costs and relatively unrationed capital compared with practically all other markets. This is a fact which conditions the demand for both consumer goods and industrial products. In the case of consumer goods, for instance, the high cost of laundresses contributes to the origins of the drip-dry shirt and the home washing machine. In the case of industrial goods, high labor cost leads to the early

---

[4] Note C. P. Kindleberger's reference to the "horizon" of the decision-maker, and the view that he can only be rational within that horizon; see his *Foreign Trade and The National Economy* (New Haven: Yale University Press, 1962), p. 15 *passim*.

[5] See, for instance, Richard M. Cyert and James G. March, *A Behavioral Theory of the Firm* (Englewood Cliffs, N.J.: Prentice-Hall, 1963), esp. Chap. 6; and Yair Aharoni, *The Foreign Investment Decision Process*, to be published by the Division of Research of the Harvard Business School, 1966.

development and use of the conveyor belt, the fork-lift truck and the automatic control system. It seems to follow that wherever there was a chance successfully to sell a new product responsive to the need to conserve labor, this chance would be apparent first to those in a position to observe the United States market.

Assume, then, that entrepreneurs in the United States are first aware of opportunities to satisfy new wants associated with high income levels or high unit labor costs. Assume further that the evidence of an unfilled need and the hope of some kind of monopoly windfall for the early starter both are sufficiently strong to justify the initial investment that is usually involved in converting an abstract idea into a marketable product. Here we have a reason for expecting a consistently higher rate of expenditure on product development to be undertaken by United States producers than by producers in other countries, at least in lines which promise to substitute capital for labor or which promise to satisfy high-income wants. Therefore, if United States firms spend more than their foreign counterparts on new product development (often misleadingly labeled "research"), this may be due not to some obscure sociological drive for innovation but to more effective communication between the potential market and the potential supplier of the market. This sort of explanation is consistent with the pioneer appearance in the United States (conflicting claims of the Soviet Union notwithstanding) of the sewing machine, the typewriter, the tractor, etc.

At this point in the exposition, it is important once more to emphasize that the discussion so far relates only to innovation in certain kinds of products, namely to those associated with high income and those which substitute capital for labor. Our hypothesis says nothing about industrial innovation in general; this is a larger subject than we have tackled here. There are very few countries that have failed to introduce at least a few products; and there are some, such as Germany and Japan, which have been responsible for a considerable number of such introductions. Germany's outstanding successes in the development and use of plastics may have been due, for instance, to a traditional concern with her lack of a raw materials base, and a recognition that a market might exist in Germany for synthetic substitutes.[6]

---

[6] See two excellent studies: C. Freeman, "The Plastics Industry: A Comparative Study of Research and Innovation," in *National Institute Economic Review*, No. 26 (Nov. 1963), p. 22 *et seq.*; G. C. Hufbauer, *Synthetic Materials and the Theory of International Trade* (London: Gerald Duckworth, 1965). A number of links in the Hufbauer arguments are remarkably similar to

Our hypothesis asserts that United States producers are likely to be the first to spy an opportunity for high-income or labor-saving new products.[7] But it goes on to assert that the first producing facilities for such products will be located in the United States. This is not a self-evident proposition. Under the calculus of least cost, production need not automatically take place at a location close to the market, unless the product can be produced and delivered from that location at lowest cost. Besides, now that most major United States companies control facilities situated in one or more locations outside of the United States, the possibility of considering a non-United States location is even more plausible than it might once have been.

Of course, if prospective producers were to make their locational choices on the basis of least-cost considerations, the United States would not always be ruled out. The costs of international transport and United States import duties, for instance, might be so high as to argue for such a location. My guess is, however, that the early producers of a new product intended for the United States market are attracted to a United States location by forces which are far stronger than relative factor-cost and transport considerations. For the reasoning on this point, one has to take a long detour away from comparative cost analysis into areas which fall under the rubrics of communication and external economies.

By now, a considerable amount of empirical work has been done on the factors affecting the location of industry.[8] Many of these studies try to explain observed locational patterns in conventional cost-minimizing terms, by implicit or explicit reference to labor cost and transportation cost. But some explicitly introduce problems of communication and external economies as powerful locational forces. These factors were given special emphasis in the analyses which were a part of the New York Metropolitan Region Study of the 1950's. At the risk of oversimplifying, I shall try to summarize what these studies suggested.[9]

---

some in this paper; but he was not aware of my writings nor I of his until after both had been completed.

[7] There is a kind of first-cousin relationship between this simple notion and the "entrained want" concept defined by H. G. Barnett in *Innovation: The Basis of Cultural Change* (New York: McGraw-Hill, 1953) p. 148. Albert O. Hirschman, *The Strategy of Economic Development* (New Haven: Yale University Press, 1958), p. 68, also finds the concept helpful in his effort to explain certain aspects of economic development.

[8] For a summary of such work, together with a useful bibliography, see John Meyer, "Regional Economics: A Survey," in the *American Economic Review*, LIII (Mar. 1963), 10–54.

[9] The points that follow are dealt with at length in the following publications: Raymond Vernon, *Metropolis, 1985* (Cambridge: Harvard University Press, 1960), pp. 38–85; Max Hall (ed.), *Made in New York* (Cambridge: Harvard University Press, 1959), pp. 3–18, 19 *passim*; Robert M. Lichtenberg, *One-Tenth of a Nation* (Cambridge: Harvard University Press, 1960), pp. 31–70.

In the early stages of introduction of a new product, producers were usually confronted with a number of critical, albeit transitory, conditions. For one thing, the product itself may be quite unstandardized for a time; its inputs, its processing, and its final specifications may cover a wide range. Contrast the great variety of automobiles produced and marketed before 1910 with the thoroughly standardized product of the 1930's, or the variegated radio designs of the 1920's with the uniform models of the 1930's. The unstandardized nature of the design at this early stage carries with it a number of locational implications.

First, producers at this stage are particularly concerned with the degree of freedom they have in changing their inputs. Of course, the cost of the inputs is also relevant. But as long as the nature of these inputs cannot be fixed in advance with assurance, the calculation of cost must take into account the general need for flexibility in any locational choice.[1]

Second, the price elasticity of demand for the output of individual firms is comparatively low. This follows from the high degree of production differentiation, or the existence of monopoly in the early stages.[2] One result is, of course, that small cost differences count less in the calculations of the entrepreneur than they are likely to count later on.

Third, the need for swift and effective communication on the part of the producer with customers, suppliers, and even competitors is especially high at this stage. This is a corollary of the fact that a considerable amount of uncertainty remains regarding the ultimate dimensions of the market, the efforts of rivals to preempt that market, the specifications of the inputs needed for production, and the specifications of the products likely to be most successful in the effort.

All of these considerations tend to argue for a location in which communication between the market and the executives directly concerned with the new product is swift and easy, and in which a wide variety of potential types of input that might be needed by the production unit are easily come by. In brief, the producer who sees a market for some new product in the United States may be led to select a United States location for production on the basis of national locational considerations which extend well beyond simple factor cost analysis plus transport considerations.

---

[1] This is, of course, a familiar point elaborated in George F. Stigler, "Production and Distribution in the Short Run," *Journal of Political Economy*, XLVII (June 1939), 305, *et seq.*

[2] Hufbauer, *op. cit.*, suggests that the low price elasticity of demand in the first stage may be due simply to the fact that the first market may be a "captive market" unresponsive to price changes; but that later, in order to expand the use of the new product, other markets may be brought in which are more price responsive.

THE MATURING PRODUCT[3]

As the demand for a product expands, a certain degree of standardization usually takes place. This is not to say that efforts at product differentiation come to an end. On the contrary; such efforts may even intensify, as competitors try to avoid the full brunt of price competition. Moreover, variety may appear as a result of specialization. Radios, for instance, ultimately acquired such specialized forms as clock radios, automobile radios, portable radios, and so on. Nevertheless, though the subcategories may multiply and the efforts at product differentiation increase, a growing acceptance of certain general standards seems to be typical.

Once again, the change has locational implications. First of all, the need for flexibility declines. A commitment to some set of product standards opens up technical possibilities for achieving economies of scale through mass output, and encourages long-term commitments to some given process and some fixed set of facilities. Second, concern about production cost begins to take the place of concern about product characteristics. Even if increased price competition is not yet present, the reduction of the uncertainties surrounding the operation enhances the usefulness of cost projections and increases the attention devoted to cost.

The empirical studies to which I referred earlier suggest that, at this stage in an industry's development, there is likely to be considerable shift in the location of production facilities at least as far as internal United States locations are concerned. The empirical materials on international locational shifts simply have not yet been analyzed sufficiently to tell us very much. A little speculation, however, indicates some hypotheses worth testing.

Picture an industry engaged in the manufacture of the high-income or labor-saving products that are the focus of our discussion. Assume that the industry has begun to settle down in the United States to some degree of large-scale production. Although the first

mass market may be located in the United States, some demand for the product begins almost at once to appear elsewhere. For instance, although heavy fork-lift trucks in general may have a comparatively small market in Spain because of the relative cheapness of unskilled labor in that country, some limited demand for the product will appear there almost as soon as the existence of the product is known.

If the product has a high income elasticity of demand or if it is a satisfactory substitute for high-cost labor, the demand in time will begin to grow quite rapidly in relatively advanced countries such as those of Western Europe. Once the market expands in such an advanced country, entrepreneurs will begin to ask themselves whether the time has come to take the risk of setting up a local producing facility.[4]

How long does it take to reach this stage? An adequate answer must surely be a complex one. Producers located in the United States, weighing the wisdom of setting up a new production facility in the importing country, will feel obliged to balance a number of complex considerations. As long as the marginal production cost plus the transport cost of the goods exported from the United States is lower than the average cost of prospective production in the market of import, United States producers will presumably prefer to avoid an investment. But that calculation depends on the producer's ability to project the cost of production in a market in which factor costs and the appropriate technology differ from those at home.

Now and again, the locational force which determined some particular overseas investment is so simple and so powerful that one has little difficulty in identifying it. Otis Elevator's early proliferation of production facilities abroad was quite patently a function of the high cost of shipping assembled elevator cabins to distant locations and the limited scale advantages involved in manufacturing elevator cabins at a single location.[5] Singer's decision to invest in Scotland as early as 1867 was also based on considerations of a sort sympathetic with our hypothesis.[6] It is not unlikely that the overseas demand for its highly standardized product was already sufficiently large at that time to exhaust the obvious scale advantages of manufacturing in a single location, especially if that location was one of high labor cost.

---

[3] Both Hirsch, *op. cit.*, and Freeman, *op. cit.*, make use of a three-stage product classification of the sort used here.

[4] M. V. Posner, "International Trade and Technical Change," *Oxford Economic Papers*, Vol. 13 (Oct. 1961), p. 323, *et seq.* presents a stimulating model purporting to explain such familiar trade phenomena as the exchange of machine tools between the United Kingdom and Germany. In the process he offers some particularly helpful notions concerning the size of the "imitation lag" in the responses of competing nations.

[5] Dudley M. Phelps, *Migration of Industry to South America* (New York: McGraw-Hill, 1963), p. 4.

[6] John H. Dunning, *American Investment in British Manufacturing Industry* (London: George Allen & Unwin, 1958), p. 18. The Dunning book

In an area as complex and "imperfect" as international trade and investment, however, one ought not anticipate that any hypothesis will have more than a limited explanatory power. United States airplane manufacturers surely respond to many "noneconomic" locational forces, such as the desire to play safe in problems of military security. Producers in the United States who have a protected patent position overseas presumably take that fact into account in deciding whether or when to produce abroad. And other producers often are motivated by considerations too complex to reconstruct readily, such as the fortuitous timing of a threat of new competition in the country of import, the level of tariff protection anticipated for the future, the political situation in the country of prospective investment and so on.

We arrive, then, at the stage at which United States producers have come around to the establishment of production units in the advanced countries. Now a new group of forces are set in train. In an idealized form, Figure I suggests what may be anticipated next.

As far as individual United States producers are concerned, the local markets thenceforth will be filled from local production units set up abroad. Once these facilities are in operation, however, more ambitious possibilities for their use may be suggested. When comparing a United States producing facility and a facility in another advanced country, the obvious production-cost differences between the rival producing areas are usually differences due to scale and differences due to labor costs. If the producer is an international firm with producing locations in several countries, its costs of financing capital at the different locations may not be sufficiently different to matter very much. If economies of scale are being fully exploited, the principal differences between any two locations are likely to be labor costs.[7] Accordingly, it may prove wise for the international firm to begin servicing third-country markets from the new location. And if labor cost differences are large enough to offset transport

---

is filled with observations that lend casual support to the main hypotheses of this paper.

[7] Note the interesting finding of Mordecai Kreinin in his "The Leontief Scarce-Factor Paradox," *The American Economic Review*, LV (Mar. 1965), 131–39. Kreinin finds that the higher cost of labor in the United States is not explained by a higher rate of labor productivity in this country.

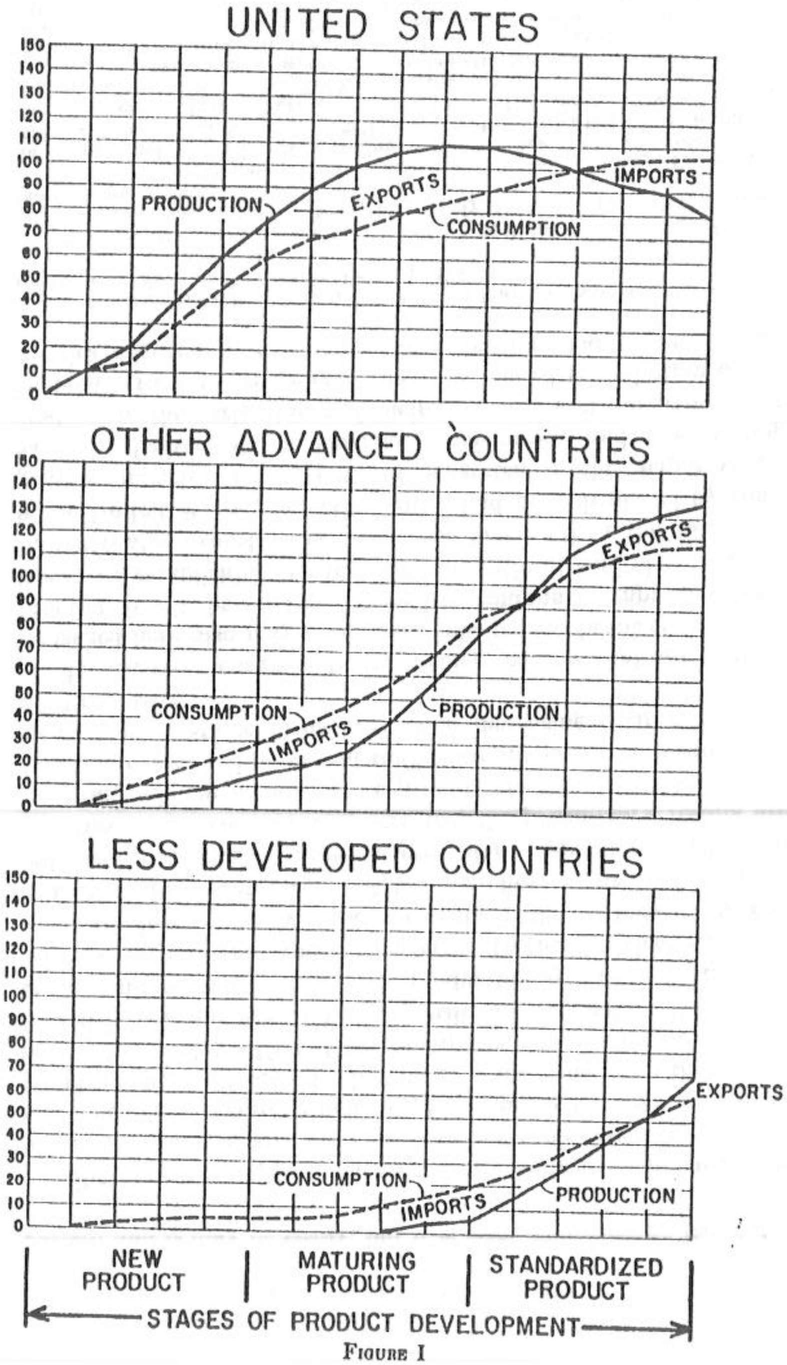


FIGURE I

costs, then exports back to the United States may become a possibility as well.

Any hypotheses based on the assumption that the United States entrepreneur will react rationally when offered the possibility of a lower-cost location abroad is, of course, somewhat suspect. The decision-making sequence that is used in connection with international investments, according to various empirical studies, is not a model of the rational process.[8] But there is one theme that emerges again and again in such studies. Any threat to the established position of an enterprise is a powerful galvanizing force to action; in fact, if I interpret the empirical work correctly, threat in general is a more reliable stimulus to action than opportunity is likely to be.

In the international investment field, threats appear in various forms once a large-scale export business in manufactured products has developed. Local entrepreneurs located in the countries which are the targets of these exports grow restive at the opportunities they are missing. Local governments concerned with generating employment or promoting growth or balancing their trade accounts begin thinking of ways and means to replace the imports. An international investment by the exporter, therefore, becomes a prudent means of forestalling the loss of a market. In this case, the yield on the investment is seen largely as the avoidance of a loss of income to the system.

The notion that a threat to the status quo is a powerful galvanizing force for international investment also seems to explain what happens after the initial investment. Once such an investment is made by a United States producer, other major producers in the United States sometimes see it as a threat to the status quo. They see themselves as losing position relative to the investing company, with vague intimations of further losses to come. Their "share of the market" is imperiled, viewing "share of the market" in global terms. At the same time, their ability to estimate the production-cost structure of their competitors, operating far away in an unfamiliar foreign area, is impaired; this is a particularly unsettling state because it conjures up the possibility of a return flow of products to the United States and a new source of price competition, based on cost differences of unknown magnitude. The uncertainty can be reduced by emulating the pathfinding investor and by investing in the same area; this may not be an optimizing investment pattern and it may be costly, but it is least disturbing to the status quo.

Pieces of this hypothetical pattern are subject to empirical tests of a sort. So far, at any rate, the empirical tests have been reassuring. The office machinery industry, for instance, has seen repeatedly the phenomenon of the introduction of a new product in the United States, followed by United States exports,[9] followed still later by United States imports. (We have still to test whether the timing of the commencement of overseas production by United States subsidiaries fits into the expected pattern.) In the electrical and electronic products industry, those elements in the pattern which can be measured show up nicely.[1] A broader effort is now under way to test the United States trade patterns of a group of products with high income elasticities; and, here too, the preliminary results are encouraging.[2] On a much more general basis, it is reassuring for our hypotheses to observe that the foreign manufacturing subsidiaries of United States firms have been increasing their exports to third countries.

It will have occurred to the reader by now that the pattern envisaged here also may shed some light on the Leontief paradox.[3] Leontief, it will be recalled, seemed to confound comparative cost theory by establishing the fact that the ratio of capital to labor in United States exports was lower, not higher, than the like ratio in the United States production which had been displaced by competitive imports. The hypothesis suggested in this paper would have the United States exporting high-income and labor-saving products in the early stages of their existence, and importing them later on.[4] In the early stages, the value-added contribution of industries engaged in producing these items probably contains an unusually high proportion of labor cost. This is not so much because the labor is particularly skilled, as is so often suggested. More likely, it is due to a quite different phenomenon. At this stage, the standardization of the manufacturing process has not gotten very far; that is to come later, when the volume of output is high enough and the degree of uncertainty low enough to justify investment in relatively inflexible, capital-intensive facilities. As a result, the production process relies relatively heavily on labor inputs at a time when the United States commands an export position; and the process relies more heavily on capital at a time when imports become important.

---

[8] Aharoni, *op. cit.*, provides an excellent summary and exhaustive bibliography of the evidence on this point.

[9] Reported in U.S. Senate, Interstate and Foreign Commerce Committee, *Hearings on Foreign Commerce*, 1960, pp. 130–39.

[1] See Hirsch, *op. cit.*

[2] These are to appear in a forthcoming doctoral thesis at the Harvard Business School by Louis T. Wells, tentatively entitled" International Trade and Business Policy."

[3] See Wassily Leontief, "Domestic Production and Foreign Trade: The American Capital Position Re-examined," *Proceedings of the American Philosophical Society*, Vol. 97 (Sept. 1953), and "Factor Proportions and the Structure of American Trade: Further Theoretical and Empirical Analysis," *Review of Economics and Statistics*, XXXVIII (Nov. 1956).

[4] Of course, if there were some systematic trend in the inputs of new products — for example, if the new products which appeared in the 1960's were more capital-intensive than the new products which appeared in the 1950's — then the tendencies suggested by our hypotheses might be swamped by such a trend. As long as we do not posit offsetting systematic patterns of this sort, however, the Leontief findings and the hypotheses offered here seem consistent.

This, of course, is an hypothesis which has not yet been subjected to any really rigorous test. But it does open up a line of inquiry into the structure of United States trade which is well worth pursuing.

The Standardized Product

Figure I, the reader will have observed, carries a panel which suggests that, at an advanced stage in the standardization of some products, the less-developed countries may offer competitive advantages as a production location.

This is a bold projection, which seems on first blush to be wholly at variance with the Heckscher-Ohlin theorem. According to that theorem, one presumably ought to anticipate that the exports of the less-developed countries would tend to be relatively labor-intensive products.

One of the difficulties with the theorem, however, is that it leaves marketing considerations out of account. One reason for the omission is evident. As long as knowledge is regarded as a free good, instantaneously available, and as long as individual producers are regarded as atomistic contributors to the total supply, marketing problems cannot be expected to find much of a place in economic theory. In projecting the patterns of export from less-developed areas, however, we cannot afford to disregard the fact that information comes at a cost; and that entrepreneurs are not readily disposed to pay the price of investigating overseas markets of unknown dimensions and unknown promise. Neither are they eager to venture into situations which they know will demand a constant flow of reliable marketing information from remote sources.

If we can assume that highly standardized products tend to have a well-articulated, easily accessible international market and

to sell largely on the basis of price (an assumption inherent in the definition), then it follows that such products will not pose the problem of market information quite so acutely for the less-developed countries. This establishes a necessary if not a sufficient condition for investment in such industries.

Of course, foreign investors seeking an optimum location for a captive facility may not have to concern themselves too much with questions of market information; presumably, they are thoroughly familiar with the marketing end of the business and are looking for a low-cost captive source of supply. In that case, the low cost of labor may be the initial attraction drawing the investor to less-developed areas. But other limitations in such areas, according to our hypothesis, will bias such captive operations toward the production of standardized items. The reasons in this case turn on the part played in the production process by external economies. Manufacturing processes which receive significant inputs from the local economy, such as skilled labor, repairmen, reliable power, spare parts, industrial materials processed according to exacting specification, and so on, are less appropriate to the less-developed areas than those that do not have such requirements. Unhappily, most industrial processes require one or another ingredient of this difficult sort. My guess is, however, that the industries which produce a standardized product are in the best position to avoid the problem, by producing on a vertically-integrated self-sustaining basis.

In speculating about future industrial exports from the less-developed areas, therefore, we are led to think of products with a fairly clear-cut set of economic characteristics.[5] Their production function is such as to require significant inputs of labor; otherwise there is no reason to expect a lower production cost in less-developed countries. At the same time, they are products with a high price elasticity of demand for the output of individual firms; otherwise, there is no strong incentive to take the risks of pioneering with production in a new area. In addition, products whose production process did not rely heavily upon external economies would be more obvious candidates than those which required a more elaborate industrial environment. The implications of remoteness also would be critical; products which could be precisely described by standardized specifications and which could be produced for inventory without fear of obsolescence would be more relevant than those which had less precise specifications and which could not easily be ordered from remote locations. Moreover, high-value items capable of absorbing significant freight costs would be more likely to appear than bulky items low in value by weight. Standardized textile products are, of course, the illustration par excellence of the sort of product that meets the criteria. But other products come to mind such as crude steel, simple fertilizers, newsprint, and so on.

---

[5] The concepts sketched out here are presented in more detail in my "Problems and Prospects in the Export of Manufactured Products from the Less-developed Countries," U.N. Conference on Trade and Development, Dec. 16, 1963 (mimeo.).

Speculation of this sort draws some support from various interregional experiences in industrial location. In the United States, for example, the "export" industries which moved to the low-wage south in search of lower costs tended to be industries which had no great need for a sophisticated industrial environment and which produced fairly standardized products. In the textile industry, it was the grey goods, cotton sheetings and men's shirt plants that went south; producers of high-style dresses or other unstandardized items were far more reluctant to move. In the electronics industry, it was the mass producers of tubes, resistors and other standardized high-volume components that showed the greatest disposition to move south; custom-built and research-oriented production remained closer to markets and to the main industrial complexes. A similar pattern could be discerned in printing and in chemicals production.[6]

In other countries, a like pattern is suggested by the impressionistic evidence. The underdeveloped south of Italy and the laggard north of Britain and Ireland both seem to be attracting industry with standardized output and self-sufficient process.[7]

Once we begin to look for relevant evidence of such investment patterns in the less-developed countries proper, however, only the barest shreds of corroboratory information can be found. One would have difficulty in thinking of many cases in which manufacturers of standardized products in the more advanced countries had made significant investments in the less-developed countries with a view of exporting such products from those countries. To be sure, other types of foreign investment are not uncommon in the less-developed countries, such as investments in import-replacing industries which were made in the face of a threat of import restriction. But there are only a few export-oriented cases similar to that of Taiwan's foreign-owned electronics plants and Argentina's new producing facility, set up to manufacture and export standard sorting equipment for computers.

If we look to foreign trade patterns, rather than foreign investment patterns, to learn something about the competitive advantage of the less-developed countries, the possibility that they are an attractive locus for the output of standardized products gains slightly more support. The Taiwanese and Japanese trade performances are perhaps the most telling ones in support of the projected pattern; both countries have managed to develop significant overseas markets for standardized manufactured products. According to one major study of the subject (a study stimulated by the Leontief paradox), Japanese exports are more capital-intensive than is the Japanese production which is displaced by imports;[8] this is what one might expect if the hypothetical patterns suggested by Figure I were operational. Apart from these cases, however, all that one sees are a few provocative successes such as some sporadic sales of newsprint from Pakistan, the successful export of sewing machines from India, and so on. Even in these cases, one cannot be sure that they are consistent with the hypothesis unless he has done a good deal more empirical investigation.

The reason why so few revelant cases come to mind may be that the process has not yet advanced far enough. Or it may be that such factors as extensive export constraints and overvalued exchange rates are combining to prevent the investment and exports that otherwise would occur.

If there is one respect in which this discussion may deviate from classical expectations, it is in the view that the overall scarcity of capital in the less-developed countries will not prevent investment in facilities for the production of standardized products.

There are two reasons why capital costs may not prove a barrier to such investment.

First, according to our hypotheses, the investment will occur in industries which require some significant labor inputs in the production process; but they will be concentrated in that subsector of the industry which produces highly standardized products capable of self-contained production establishments. The net of these specifications is indeterminate so far as capital-intensiveness is concerned. A standardized textile item may be more or less capital-intensive than a plant for unstandardized petro-chemicals.

---

[6] This conclusion derives largely from the industry studies conducted in connection with the New York Metropolitan Region study. There have been some excellent more general analyses of shifts in industrial location among the regions of the United States. See e.g., Victor R. Fuchs, *Changes in the Location of Manufacturing in the United States Since 1929* (New Haven: Yale University Press, 1962). Unfortunately, however, none has been designed, so far as I know, to test hypotheses relating locational shifts to product characteristics such as price elasticity of demand and degree of standardization.

[7] This statement, too, is based on only impressionistic materials. Among the more suggestive, illustrative of the best of the available evidence, see J. N. Toothill, *Inquiry into the Scottish Economy* (Edinburgh: Scottish Council, 1962).

[8] M. Tatemoto and S. Ichimura, "Factor Proportions and Foreign Trade: the Case of Japan," *Review of Economics and Statistics*, XLI (Nov. 1959), 442–46.

Besides, even if the capital requirements for a particular plant are heavy, the cost of the capital need not prove a bar. The assumption that capital costs come high in the less-developed countries requires a number of fundamental qualifications. The reality, to the extent that it is known, is more complex.

One reason for this complexity is the role played by the international investor. Producers of chemical fertilizers, when considering whether to invest in a given country, may be less concerned with the going rate for capital in that country than with their opportunity costs as they see such costs. For such investors the alternatives to be weighed are not the full range of possibilities calling for capital but only a very restricted range of alternatives, such as the possibilities offered by chemical fertilizer investment elsewhere. The relevant capital cost for a chemical fertilizer plant, therefore, may be fairly low if the investor is an international entrepreneur.

Moreover, the assumption that finance capital is scarce and that interest rates are high in a less-developed country may prove inapplicable to the class of investors who concern us here.[9] The capital markets of the less-developed countries typically consist of a series of water-tight, insulated, submarkets in which wholly different rates prevail and between which arbitrage opportunities are limited. In some countries, the going figures may vary from 5 to 40 per cent, on grounds which seem to have little relation to issuer risk or term of loan. (In some economies, where inflation is endemic, interest rates which in effect represent a negative real cost are not uncommon.)

These internal differences in interest rates may be due to a number of factors: the fact that funds generated inside the firm usually are exposed to a different yield test than external borrowings; the fact that government loans are often floated by mandatory levies on banks and other intermediaries; and the fact that funds borrowed by governments from international sources are often reloaned in domestic markets at rates which are linked closely to the international borrowing rate, however irrelevant that may be. Moreover, one has to reckon with the fact that public international lenders tend to lend at near-uniform rates, irrespective of the identity of the borrower and the going interest rate in his country. Access to capital on the part of underdeveloped countries, therefore, becomes a direct function of the country's capacity to propose plausible projects to public international lenders. If a project can plausibly be shown to "pay its own way" in balance-of-payment and output terms at "reasonable" interest rates, the largest single obstacle to obtaining capital at such rates has usually been overcome.

Accordingly, one may say that from the entrepreneur's viewpoint certain systematic and predictable "imperfections" of the capital markets may reduce or eliminate the capital-shortage handicap which is characteristic of the less-developed countries; and, further, that as a result of the reduction or elimination such countries may find themselves in a position to compete effectively in the export of certain standardized capital-intensive goods. This is not the statement of another paradox; it is not the same as to say that the capital-poor countries will develop capital-intensive economies. All we are concerned with here is a modest fraction of the industry of such countries, which in turn is a minor fraction of their total economic activity. It may be that the anomalies such industries represent are systematic enough to be included in our normal expectations regarding conditions in the less-developed countries.

* * * * *

Like the other observations which have preceded, these views about the likely patterns of exports by the less-developed countries are attempts to relax some of the constraints imposed by purer and simpler models. Here and there, the hypotheses take on plausibility because they jibe with the record of past events. But, for the most part, they are still speculative in nature, having been subjected to tests of a very low order of rigorousness. What is needed, obviously, is continued probing to determine whether the "imperfections" stressed so strongly in these pages deserve to be elevated out of the footnotes into the main text of economic theory.

HARVARD GRADUATE SCHOOL OF BUSINESS ADMINISTRATION

---

[9] See George Rosen, *Industrial Change in India* (Glencoe, Ill.: Free Press, 1958). Rosen finds that in the period studied from 1937 to 1953, "there was no serious shortage of capital for the largest firms in India." Gustav F. Papanek makes a similar finding for Pakistan for the period from 1950 to 1964 in a book about to be published.

16

LIPPINCOTT LIBRARY
DOCUMENT DELIVERY SERVICE

WHARTON REPROGRAPHICS

Environment and Planning A 1994, volume 26, pages 509-526

Restructuring of items within a product line: a case study of Japanese multinational enterprises in the electric machinery industry

T Hayashi
Faculty of Economics, Nanzan University, 18 Yamazato-cho, Showa-ku, Nagoya 466, Japan

Received 12 May 1992; in revised form 26 October 1992

Abstract. Through an examination of the intrafirm division of labour between the parent companies of Japanese multinational enterprises (MNEs) and their Asian affiliates in the electric machinery industry, it is illustrated that the restructuring of items within product lines has been the dominant pattern of restructuring behaviour. There are three characteristics to this pattern: (1) vertical restructuring, that is, upgraded items are continually innovated by Japanese parents; (2) horizontal restructuring, that is, the production of each item is transferred to Asian affiliates after a certain time lag; (3) the time lag in the transfer of the production of an item to Asian affiliates has been shortened, and Asian affiliates have achieved product upgrading at a faster pace than have the Japanese parents. These characteristics will be described in detail for two product lines: video cassette recorders and capacitors. Then some supporting evidence is presented for the third characteristic of this pattern of restructuring. Further, this third trend is explained in association with the change in cost conditions which have arisen from recent high-growth performance in the region. In other words, in the recent dynamism in the Asia-Pacific Basin, the horizontal restructuring of Japanese MNEs has tended to become faster in terms of production transfer.

1 Introduction

1.1 Restructuring by Japanese multinational enterprises in the Asia-Pacific Basin

In the past two decades, the developing countries in the Asia-Pacific Basin, most notably the Asian newly industrialised economies (NIEs) (Hong Kong, Korea, Singapore, and Taiwan) and ASEAN-4 countries (Indonesia, Malaysia, the Philippines, and Thailand) have achieved a steady and high rate of economic growth which has been accompanied by an even higher growth rate in trade, capital flow, and foreign direct investment (FDI).

The dynamism arising from this high-growth performance seems to have caused a significant change in cost conditions for manufacturing firms in this region. For instance, because of the expansion in labour-intensive manufacturing sectors, the huge surplus in unskilled labour has been absorbed and their wage level has begun to rise. At the same time, a substantial amount of skilled labour has been created and accumulated. The development of the final-product manufacturers has been followed by the development of the component manufacturers, which assures closer linkage between final and intermediate producers. In addition, the domestic market has grown significantly, along with the rise in per capita income in developing countries in the region. These changes in cost conditions are expected to cause a significant impact on manufacturing firms' restructuring behaviour with respect to production structure.

In this paper, I examine manufacturing firms' restructuring behaviour in the Asia-Pacific Basin in relation to the recent change in cost conditions. For this purpose, I focus on Japanese multinational enterprises (MNEs) and the intrafirm division of labour between the Japanese parents and their Asian affiliates.

MNEs and their intrafirm division of labour have attracted much analytical attention, as MNEs are expected to have two major advantages in pursuing restructuring behaviour. First, MNEs are endowed with their own managerial resources (so-called 'firm-specific intangible assets', such as managerial know-how and production technology). Second, they have wide locational boundaries (parents in the home country and affiliates abroad) which offer them a wider selection of production-factor prices.(1)

In fact, it is possible to observe that the two advantages shared by MNEs can explain the two major characteristics of their ongoing restructuring behaviour: (1) vertical restructuring, or the upgrading over time of products which is achieved through managerial resources, and (2) horizontal restructuring, or the coordination of production across locations between the Japanese parents and their Asian affiliates.(2)

1.2 Restructuring of items within a product line

In particular, from interviews with eighteen Japanese MNEs in the electric machinery industry, I present the restructuring of items within a product line(3) as a dominant pattern of restructuring behaviour in MNEs which has two major characteristics.

1 Within a product line, new items which are upgraded in terms of product design, general quality, and the attached characteristics are continually undergoing innovations by Japanese parents (vertical restructuring).

2 When each item passes through a certain stage of its life cycle, its production is transferred from Japanese parents to Asian affiliates (horizontal restructuring).

In addition to these two characteristics, I have observed a third characteristic to this stylised pattern, and it is associated with the recent change in cost conditions in the Asia-Pacific Basin.

If the first two characteristics described above are combined, the following proposition can be made as a hypothetical benchmark for the relative speed of vertical and horizontal restructuring.

Proposition 1: Suppose that every item is succeeded by a newer item within the same product line *after a certain time interval* (h_J) of its production initiation; that is, the product upgrading takes place at intervals of h_J in the Japanese parents. Suppose also that the production of each item within a product line is transferred to Asian affiliates *with a constant time lag* (t^*) from its production initiation; that is, the time lag in the transfer of production is constant across items. Then, with a constant time lag of t^*, the product upgrading takes place in Asian affiliates at intervals of h_A, identical to the interval in the Japanese parents: that is, $h_A = h_J$.

However, interviews show that t^* is not necessarily constant, but is likely to decrease over time. Here, in the context of recent growth performance and the resulting changes in cost conditions in the Asia-Pacific Basin, the third characteristic of the restructuring of items within a product line is proposed.(4)

---

(1) In the literature, factors associated with the firm's specific managerial resources as well as geographical location are analysed with respect to the nature of MNEs and their FDI activity. In particular, in the eclectic theory proposed by Dunning (1979), the nature of FDI activity is explained by (1) ownership advantages (firm-specific intangible assets), (2) locational advantages, and (3) internationalisation advantages.

(2) A similar idea is presented by Hara (1989), Urata (1991), MITI (1989a), and EPA (1990). They explained that, since a rapid appreciation of Japanese yen began in 1985, Japanese manufacturing firms have accelerated their adjustment in three ways: (1) by upgrading products and/or production processes in the Japanese parent companies, (2) by shifting the location of production from the Japanese parents to foreign affiliates for products and/or production processes which have lost their cost competitiveness in Japan, and (3) by establishing a more organised system of intrafirm division of labour.

(3) A product line is defined as a group of similar products consisting of various items. In this study, eight groups of products—colour televisions, video cassette recorders, facsimile machines, telephone machines (including key telephones), capacitors, resistors (including potentiometers), electric motors, and switches—are considered to be product lines.

3 The recent high-growth performance in the Asia-Pacific Basin has narrowed the gap in cost conditions, the differences in factor prices between Japanese parents and their Asian affiliates.(5) This change in the cost conditions has shortened t^*, the time lag in the transfer of production of an item. As a result, the product upgrading takes place in Asian affiliates at an interval (h_A), which is shorter than that of the Japanese parents (h_J).

In the following sections I will investigate the three characteristics of restructuring by using data from interviews with and questionnaires sent to Japanese MNEs in the electric machinery industry. In particular, in section 2, I illustrate the ongoing restructuring of items within a product line by identifying the first two characteristics, and in section 3, I discuss the third characteristic. In section 4, I summarise the findings and propose some issues for further research.

2 An illustration of the restructuring of items within a product line

2.1 Data obtained from interviews and questionnaires

The interviews and questionnaires were conducted between November 1990 and January 1991 and were sponsored by the Resource Systems Institute, East–West Center. Initially, thirty-two Japanese MNEs in the electric machinery industry which have been actively involved in FDI in Asia were chosen, and thirteen product lines which have been widely produced both by Japanese parents and their Asian affiliates were selected. The data actually obtained cover eighteen firms (six of them are final-product electronic manufacturers and twelve of them are electronic-component manufacturers) and eight product lines (four of them are final-product electronics and the other four are electronic components; see footnote 3). If one firm's response for one product line is counted as a single case, then data were obtained for twenty-eight cases (fourteen cases concern final-product electronics and the others relate to electronic components).

An overview of the data obtained is given in table 1 (over), which indicates identification letters for the eighteen firms; the serial numbers which were given to the twenty-eight cases; the product lines; whether the firm's response was through interview or from a written questionnaire; and the position of interviewees within the firms or the division which was responsible for answering the questionnaire.

2.2 Two characteristics of the restructuring of items within a product line

The first two characteristics described in section 1.2 can be identified (a) by observing evolutionary change in items within the product line, and (b) by comparing the characteristics of items produced by Asian affiliates with those of the Japanese parents. In this section, in order to illustrate the restructuring I look for these identifiers in the two product lines where the stylised pattern is seen relatively clearly [video cassette recorders (VCRs) and capacitors].

---

(4) Given that dt^* is the change in the time lag in production transfer, then $h_A = h_J + dt^*$ should generally be satisfied. Although there might be various factors affecting h_J and dt^*, I only argue that the possible reduction in h_A is caused by the declining t^* associated with the recent change in the relative cost conditions between the Japanese parents and their Asian affiliates. In this way the impact of the dynamism arising from recent growth performance in the Asia-Pacific Basin on the restructuring behaviour by Japanese MNEs can be focused upon.

(5) In particular, in terms of the factor prices of the Japanese parents relative to their Asian affiliates, initially expensive prices for skilled labour, communication, and transportation inputs are expected to be raised, tending to shorten t^*. However, the initially expensive cost of unskilled labour is expected to be lowered, tending to lengthen t^* (see section 3).

Table 1. Characteristics of the multinational enterprises studied.

Firm	Case	Product line	Data source	Position of interviewees and the sections responsible for written questionnaire
(a) Manufacturers of final-product electronics				
A	1	televisions	I	Manager, planning department, consumer products group
	7	VCRs		
B	2	televisions	Q	Public relations department
	8	VCRs		
C	3	televisions	Q/I	Public affairs department, consumer product group
	11	fax machines	I	Manager, planning office, communication system devices group
	13	telephones		
D	4	televisions	I	Manager, overseas department, manufacturing promotion headquarters
E	5	televisions	I	General manager, manufacturing department, international manufacturing strategy group
	9	VCRs		
	12	fax machines		Vice counselor, management planning division
	14	telephones		
F	6	televisions	I	Manager, public communications office
	10	VCRs		
(b) Manufacturers of electronic components				
G	23	resistors	I	Manager, administration section, overseas operations supporting department
H	19	motors	I	Director, management planning division
	26	switches		Manager, management planning division
I	20	motors	I	Public relations and advertising section, sales-manufacturing support and coordination department
J	21	motors	I	Manager, public relations section
	27	switches		
K	15	capacitors	I	General affairs department
L	16	capacitors	Q	Public relations department
M	24	resistors	Q	Public relations department
N	22	motors	I	Manager, overseas business department
O	17	capacitors	I	Vice president
P	18	capacitors	I	Manager, corporate planning department Vice section chief, corporate planning department
Q	28	switches	I	General manager, planning department Manager, sales department
R	25	resistors	I	Public relations section, general affairs department

Note: VCR, video cassette recorder; I interview; Q written questionnaire.

2.2.1 *Video cassette recorders (cases 7–10)*

Evolutionary change: Since their introduction in the home market, VCRs have been modified significantly. Quality upgrading of VCRs has occurred in terms of product design and general quality (table 2).

First, in terms of product design, portable VCRs have emerged and joined the ranks of video cameras and conventional VCRs. Second, as an improvement in general quality, the sound quality of VCRs has been enhanced through high-fidelity (hi-fi) sound systems, and the visual quality of VCRs has been enhanced through supervideo home systems (S-VHS). In addition, these new sound and visual systems have been accompanied by an increase in the number of rounding heads.

A comparison: A comparison of the product characteristics of VCRs produced by Asian affiliates and by Japanese parents fits the stylised pattern of restructuring in all four cases (table 3). First, the Asian affiliates specialise in producing conventional floor-type VCRs, whereas the Japanese parents produce portable VCRs as well (although the share of portable VCRs seems to be rising). Second, the Asian affiliates produce VCRs of standard sound and visual qualities, whereas Japanese parents produce VCRs with enhanced sound and visual qualities through hi-fi sound systems and S-VHS. In addition, whereas Asian affiliates produce VCRs with two heads only, Japanese parents produce VCRs with four, six, and eight rounding heads.

Table 2. Evolution of video cassette recorders (VCRs).

Product design

Emergence and growing share of portable VCRs

In 1983 production started; in 1987 there was a significant size reduction through compact cassette tapes.[a, b]

The share of portable VCRs in total VCR production in Japan increased: value was 28.4% in 1987, 34.8% in 1989; quantity was 14.1% in 1987, 17.6% in 1989.[b]

General quality

Sound quality has been enhanced by high-fidelity sound system since 1984.[a, b]

Share of VCRs with high-fidelity sound system in domestic shipment of total floor-type VCRs has increased from 25.5% in 1985 to 70.3% in 1989.[b]

Visual quality has been enhanced by supervideo home system since 1987.[a, b]

Quality enhancement has been achieved by an increase in the number of rounding heads: from two heads to four heads (sound and visual heads are separated) to six and eight heads (for slower mode and erasing).[c]

[a] Source: Asakura (1990). [b] Source: EIAJ (1990a; 1990b). [c] Source: case 9, interview.

Table 3. Characteristics of video cassette recorders (VCRs) produced by Asian affiliates and by Japanese parents (sources: cases 7–10, interviews and questionnaires).

Asian affiliates	Japanese parents
Product design	
Only floor-type VCRs; no portable VCRs	Floor-type and portable VCRs
General quality	
Standard sound quality; no high-fidelity sound system	Enhanced (high-fidelity) sound quality
Standard visual quality	Enhanced (supervideo home systems) visual quality
Only two rounding heads[a]	Four, six, and eight rounding heads[a]

[a] Source: case 9 only.

2.2.2 *Ceramic and aluminium electrolytic capacitors (cases 15, 16, and 18)*

Evolutionary change: As shown in table 4, the original lead-type capacitor that cannot be used with automated packaging machines has been replaced with a lead-type capacitor that can. More recently, these lead-type capacitors have been replaced with chip-type capacitors which are compatible with surface mounting technology (SMT).(6) At the same time, product diversification has taken place in capacitors, and composite products such as hybrid integrated circuits (ICs)(7) have emerged. Another notable recent change is the reduction in the size of the chip-type capacitors. This improvement is essential in the size reduction of final electronic products such as VCRs and portable cassette tape players.

---

(6) SMT is a recently developed technology in the package board (PCB) assembling process. In the PCB assembling process, the elements, such as capacitors, resistors, and transistors, are packed onto the PCB. In conventional packaging technology, each element is attached with lead line, and the lead line is inserted into a minute hole in the PCB and is fixed with solder. However, in the packaging of SMT, the chip-type elements, which are without lead lines, are attached directly to the PCB, which is without holes, and are fixed with solder. In this way, more elements can be simultaneously attached to a PCB at a higher speed, and the soldering process is simpler.

A comparison: Analysis of the data indicates that both for ceramic capacitors (cases 15 and 18) and for aluminium electrolytic capacitors (case 16), the stylised pattern of intrafirm transfer of production holds true (see table 5).

In the case of ceramic capacitors [table 5(a)], the Asian affiliates produce conventional lead-type capacitors which still make up a large share of the total production by the affiliates. Asian affiliates also produce the more recently developed chip-type capacitors, the majority of which are, however, large in size. Japanese parents produce conventional lead-type capacitors, but these capacitors are for specialised use (for example, high-voltage durable capacitors and capacitors with a high electrostatic capacity). Nevertheless, Japanese parents mainly produce the more recently developed chip-type capacitors, and the size of these is relatively small. In addition, Japanese parents produce hybrid ICs, whereas the Asian affiliates do not yet do so.

Table 4. Evolution of ceramic and aluminium electrolytic capacitors.

Product design of ceramic capacitors evolved: from lead-type unusable in automated packaging, to lead-type usable in automated packaging, to chip-type and composite products (hybrid integrated circuits, etc).[a, b]

The chip-type capacitor (ceramic and aluminium electrolytic) has undergone a significant reduction in size: from 3.2 mm × 1.6 mm in 1977, to 1.6 mm × 0.8 mm in 1985, to 1.0 mm × 0.5 mm in 1989.[b, c]

[a] Source: case 18, interviews. [b] Source: Shimura (1986). [c] Source: EIAJ (1990a).

Table 5. Characteristics of capacitors produced by Asian affiliates and by Japanese parents.

Asian affiliates	Japanese parents
(a) Ceramic capacitors	
Lead-type occupies relatively high share of production[a]	Lead-type occupies lower share of production and is produced for specialised use.
Chip-type of larger sizes, occupies relatively high share of production:[a, c] minimum size is 1.6 mm × 0.8 mm[a]	Chip-type of smaller sizes, occupies higher share of production:[a, c] minimum size is 1.0 mm × 0.5 mm[a]
Hybrid integrated circuits are not yet produced	Hybrid integrated circuits are produced[c]
(b) Aluminium electrolytic capacitors[b]	
Lead-type capacitors for generic use are produced	Lead-type capacitors for generic and specialised use are produced
Chip-type capacitors are not yet produced	Chip-type capacitors are produced and their share of production is rising

[a] Source: case 15, interview. [b] Source: case 16, questionnaire. [c] Source: case 18, interview.

---

(7) The hybrid IC includes the capacitors, as well as resistors, monolithic ICs, etc, which form an IC on a board, which is called an insulator substrate. By integrating several components into one unit, the number of lead lines is reduced, lowering the defect ratio. In addition, various customised circuits with a higher density can be flexibly designed with relatively small modification costs.

In terms of aluminium electrolytic capacitors [table 5(b)], Asian affiliates specialise in conventional lead-type capacitors for generic use, and recently developed chip-type capacitors are not yet produced. Japanese parents, however, produce for both generic and specialised use. They also produce recently developed chip-type capacitors, and the production share of these is growing.

2.2.3 *Capacitors made with a certain mineral dielectric (case 17)*

Evolutionary change in other capacitors: Capacitors in which a certain mineral dielectric is used have experienced product diversification with the development of composite products that can be used in the capacitors[8]—particularly L/C filters, in which coils and capacitors are integrated, and hybrid ICs.

A comparison: Production of these types of capacitors have followed the stylised pattern of intrafirm transfer of production, and the transfer lag seems to have declined in recent years (table 6).

The production of conventional capacitors by the Asian affiliates began in 1974, following a seventeen-year lag from when production first began in the Japanese parent companies. The time lag for the transfer of L/C filters from Japanese parents to the Asian affiliates, where production began in 1982, was eleven years. In the case of more recently developed hybrid ICs, it has taken only five years from the time the product was first produced by Japanese parents to the time when production was transferred to Asian affiliates (in 1987).

Thus, Asian affiliates and Japanese parents currently produce all three products (capacitors, L/C filters, and hybrid ICs). However, Japanese parents specialise in the production of items that are for a very particular use; typically, these products are for newly developed high-technology products such as equipment for an integrated services digital network (ISDN), laser application equipment, and high-vision televisions. In contrast, Asian affiliates produce items of standard quality for generic use.

Table 6. Characteristics of capacitors of a certain mineral dielectric (source: case 17, interview).

Asian affiliates	Japanese parents
Various kinds of capacitors and related products made: capacitors, 1975 onwards L/C filters, 1982 onwards hybrid integrated circuits, 1987 onwards	Various kinds of capacitors and related products made: capacitors, 1957 onwards L/C filters, 1971 onwards hybrid integrated circuits, 1982 onwards
Products are oriented for generic use	Products are oriented to specialised use, typically for newly developed products: integrated services digital network equipment, laser application equipment

2.3 Summary of survey results

In all eight product lines, the two characteristics of restructuring have been identified in most of the twenty-eight cases. As shown in tables 7 and 8, this stylised pattern of restructuring is identified in eleven of the fourteen cases of final-product electronics and in thirteen of the fourteen cases of electronic components.

---

(8) There has been a move to composite capacitor-using products: from a basic capacitor in 1957, to use of L/C filters (coils and capacitors are integrated) from 1971, to the hybrid integrated circuit from 1982 (case 17, interview).

The cases where the stylised pattern is not clearly identified warrant some brief comments. For the fax-machine product line (cases 11 and 12), comparison of products currently produced by Asian affiliates and Japanese parents suggest that the stylised pattern may be taking place. However, the fax machine has experienced substantial market diversification, which makes it difficult to judge whether or not the stylised pattern has actually taken place. Similarly, in the case of the key-telephone product line (case 13), the existence of the stylised pattern was suggested, but there was insufficient information available on the evolution of the product to

Table 7. Results: identification of stylised patterns.

Firm	Case number	Product line	Pattern identified?
(a) Final-product electronics			
A	1	televisions	yes
B	2	televisions	yes
C	3	televisions	yes
D	4	televisions	yes
E	5	televisions	yes
F	6	televisions	yes
A	7	video cassette recorders	yes
B	8	video cassette recorders	yes
E	9	video cassette recorders	yes
F	10	video cassette recorders	yes
C	11	facsimile machines	not clearly
E	12	facsimile machines	not clearly
C	13	key telephones	not clearly
E	14	telephones	yes
(b) Electronic components			
K	15	capacitors	yes
L	16	capacitors	yes
O	17	capacitors	yes
P	18	capacitors	yes
H	19	motors	yes
I	20	motors	yes
J	21	motors	yes
N	22	motors	yes
G	23	resistors	not clearly
M	24	resistors	yes
R	25	resistors	yes
H	26	switches	yes
J	27	switches	yes
Q	28	switches	yes

Table 8. Number of cases where the stylised pattern is identified.

Product line	Number of cases	Pattern identified	Product line	Number of cases	Pattern identified
(a) Final-product electronics			*(b) Electronic components*		
Televisions	6	6	Capacitors	4	4
Video cassette recorders	4	4	Motors	4	4
Facsimile machines	2	0	Resistors	3	2
Telephones	2	1	Switches	3	3
Total	14	11	Total	14	13

make a final determination. However, in the case of the potentiometer product line (case 23), the distinction between products made by Asian affiliates and Japanese parents was not clear, although the pattern of intrafirm division of labour appeared generally to be determined by the local demand for the product in each production location.

3 The shortening time lag in the production transfer of items

3.1 Supporting evidence for a shortening time lag

Although sufficient evidence is not yet available, the data obtained from interviews and questionnaires generally support a recent trend towards a shortening time lag in the production transfer of an item to Asian affiliates.

Table 9 provides a summary of the responses to the first question asked in the interviews and questionnaires: "In what year did your company start producing the product in Japanese parents, in affiliates in the Asian NIEs, and in the ASEAN-4 countries?". First, the time lag for production transfer from the Japanese parents to their Asian affiliates for each product line is calculated. This is done in two ways: the time lag for a particular MNE which was first to begin production in the Japanese parent is taken; and the average time lag for all MNEs is also calculated.

Table 9. The time lag in the intrafirm transfer of production from Japanese parents to Asian affiliates (source: author's survey).

Product line	First case[a]		Average[b]		
	start[c]	time lag	no. of cases	start[c]	time lag
(a) Final-product electronics					
Televisions	1952	17	6	1952.8	16.2
Telephones	1955	32	2	1971.0	17.0
Facsimile machines	1972	15	2	1977.0	11.0
Video cassette recorders	1975	15	4	1976.8	13.0
(b) Electronic components					
Capacitors	1931	41	4	1947.3	24.0
Resistors	1944	21	3	1950.0	19.0
Switches	1953	16	3	1957.3	16.0
Motors	1958	6	3	1960.7	15.0

[a] Time lag calculated by taking the multinational enterprise which first started the product line in the Japanese parent.

[b] Time lag calculated by considering the average of all multinational enterprises with the specified product line.

[c] The year production began in Japan.

Strictly speaking, these values are not accurate measures of the time lag as described in section 1.2; that is, the time lag was specified as the lag for each item within the product line, but the values in these tables indicate the lag for each product line. Nonetheless, it is quite insightful to find a general trend which parallels the third assertion that the newer the product line, the shorter the time lag in the transfer of production.

Second, although a precise examination of the time lag for each item is difficult, some cases of product lines of colour televisions (cases 4 and 5), capacitors (case 17), electronic motors (case 20), and fixed resistors (case 24) examined in the interviews and questionnaires actually revealed that the lag is declining for each item as well as for each product line. For instance, table 6 (case 17) indicates this trend for capacitors with certain mineral dielectrics. As another extreme instance, case 20 illustrates that the Japanese parent of an electronic-motor manufacturer is now specialising in research and development (R&D) activities, whereas its Asian affiliates are in charge of producing 99% of the total output in terms of quantity.

3.2 A shortened time lag in the high-growth performance of the Asia-Pacific Basin

As shown above, the results of interviews and questionnaires generally indicated that the time lag in the production transfer of items has recently been shortened. In other words, the production of each item tends to be transferred from the Japanese parents to their Asian affiliates at an earlier stage in the life cycle of an item, and hence Asian affiliates have upgraded their products at a faster pace than have the Japanese parents.

In the following analysis, I discuss the observed shortened time lag in the context of a dynamic change in cost conditions arising from the recent high-growth performance in the Asia-Pacific Basin. For this purpose, and using a model developed in my PhD thesis (Hayashi, 1992), I examine the determinant mechanism of the intrafirm transfer of production, and I point out several factors which affect the time lag in the production transfer of items.

3.2.1 *A determinant mechanism of intrafirm transfer of production*

Using a version of figure 1, I have in an earlier publication explained the stylised transfer of production along the life cycle of an item (Hayashi, 1992). The explanation is as follows. Suppose that MNEs choose the production location of a certain item such that they can minimise its production cost at any given time. Then, the stylised transfer from the Japanese parents to their Asian affiliates will occur if the production cost of a certain item for Japanese parents, D_J, relative to that for Asian affiliates, D_A, increases monotonically over time, t [that is, $D'(t) > 0$, where $D(t) = D_J(t)/D_A(t)$; see figure 1]. At the early stages of the life cycle of an item, the production cost for the Japanese parents is lower than that for their Asian affiliates, and therefore the item is produced by the parent. As the life cycle proceeds, the relative production cost for the Japanese parents increases. Finally, after a certain period (t^* in figure 1), its value exceeds 1 and the production is transferred to the Asian affiliates.

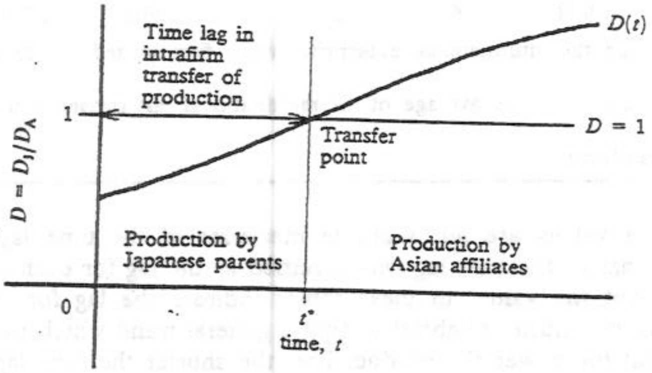

Time lag in intrafirm transfer of production
D(t)
1
D = 1
Transfer point
D = D_J/D_A
Production by Japanese parents
Production by Asian affiliates
0
t*
time, t

Figure 1. The increase in production costs for the Japanese parents, D_J, relative to those in the Asian affiliates, D_A, over the life cycle of an item.

The results of the interviews and questionnaires on the reasons why a particular item, rather than other items in the same product line, was chosen for production in the Asian affiliates (Hayashi, 1992, chapter 5) supported the model (chapter 3) in which six determinant factors were derived which led to a positive slope in the curve $D(t)$; that is, to a rise in production costs for Japanese parents relative to those for the Asian affiliates along the life cycle of an item. These six factors are as follows.

(1) As the item proceeds along its life cycle, the relative importance to the cost of unskilled labour, W, increases. Thus, the cost advantage of Asian affiliates in terms of availability of inexpensive unskilled labour becomes more significant.

(2) The relative importance of the cost of skilled labour, U, declines. Thus, the cost disadvantage in terms of the relatively high cost of skilled labour in the Asian affiliates becomes less significant.

(3) The relative importance of communication inputs, V, declines. Thus, the cost disadvantage in terms of expensive communication costs in the Asian affiliates becomes less significant.

(4) The communication costs which were initially expensive for the Asian affiliates decline because of a shift in demand and the rising positive externality between final-product manufacturers and component manufacturers.

(5) The relative importance of transportation inputs, Z, declines. Thus, the cost disadvantage arising from expensive transportation inputs becomes less significant.

(6) The transportation costs which were initially expensive decline because of a shift in demand and the rising positive externality between final-product manufacturers and component manufacturers.

Given these six determinant factors and the monotonically increasing curve of $D(t)$, one important proposition can be made about the determination of t^*.

Proposition 2: Any exogenous factor which shifts the curve $D(t)$ to the left (that is, an increase in D_J/D_A) will shorten t^*. Conversely, any exogenous factor which shifts the curve to the right (that is, a decrease) will lengthen t^* (see figure 2).

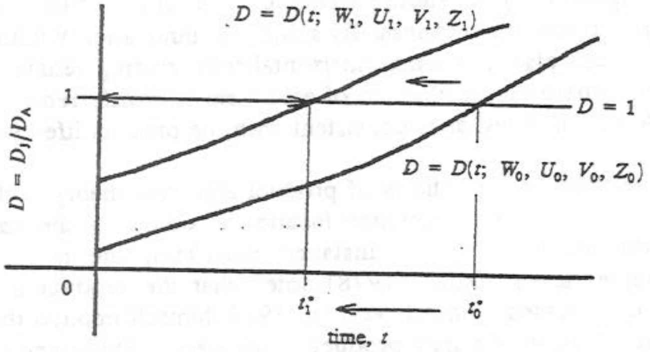


Figure 2. The shift in importance of production costs in the Japanese parents (D_J) relative to those in the Asian affiliates (D_A) and a decreasing time lag for production transfer (t^*). Note: for definitions of W, U, V, and Z, see points (1)–(6) in subsection 3.2.1 of text (subscripts denote time).

3.2.2 *A brief survey of product life-cycle theory*

It should be noted here that the determinant factors described above are closely related to those in product life-cycle theory, as asserted by Vernon (1966), Hirsch (1967), and others. Hence, the literature on product life-cycle theory as well as some criticisms of it are now briefly reviewed in association with an analysis.

The salient feature of the product life-cycle model is the assumed stylised change in technology characteristics for a certain product towards a standardised or matured product. In particular, it is asserted that the uncertainty associated with market demand, product design, and production technology for a given product is likely to decline over time, which leads to a stylised transfer in production location for the product from the United States, say, to other advanced countries, and, finally, to developing countries.

That these factors are still relevant to a significant extent in determining production location is supported by my earlier work in which I derived mathematically factors (1)–(6) which lead to a positive slope for $D = D_J/D_A$ (Hayashi, 1992) and by the increase in the production costs for the Japanese parents relative to those in the Asian affiliates. In particular, factors (3)–(6), which had been considered as 'geographical externality factors' in Hirsch (1967) but had not been distinguished from each other, were explicitly identified by the incorporation of communication and transportation inputs in the model.

In the 1960s and 1970s, a number of empirical tests were carried out, which generally validated the product life-cycle theory, including regression analysis by Keesing (1967), Gruber et al (1967), and Finger (1975), and industrial case studies by Stobaugh (1972) and Hirsch (1972). However, since the latter half of the 1970s, some questions about the empirical validity of the product life-cycle model have arisen.

The first group of questions was associated with the assumption that products 'standardise' or 'mature'. According to Walker (1979), Freeman (1982), and Abernathy et al (1983), there need not be technological stability nor standardisation of products. Rather, it is more likely that there will be an increase in the rate of technical change, or, to borrow Schumpeter's terminology, "the perennial gale of creative destruction". In fact, Walker (1979) and Baba (1989) illustrated the continual upgrading in production technologies and product designs in the chemical and television industries, for which a typical product life-cycle pattern was claimed by Stobaugh (1972).

In accordance with Walker's and Baba's observations, the results from this study have shown the continual upgrading by Japanese parents to be vertical restructuring behaviour, where new upgraded items are continually subject to innovation within a product line. However, this study also shows that horizontal restructuring behaviour takes place at the same time, where the production of each item is transferred from Japanese parents to their Asian affiliates, as is consistent with the product life-cycle theory.

The second group of questions of the validity of product life-cycle theory is the assumption of a stylised order of shift in production location as well as of the time lag in the transfer of production location. For instance, illustrating the cases of digital watches and disposable razors, Giddy (1978) noted that the product life-cycle pattern can be highly compressed. Indeed, Vernon (1979) himself implied that the time lag between the introduction of a new product in the United States and its first production in a foreign location has been decreasing rapidly.

Vernon (1979) pointed out the two possible reasons for the shortened time lag in intrafirm transfer of production. The first is the growth and the spread of MNEs in forming their global network systems, because these reduce the communication cost and hence reduce the importance of geographical proximity among markets, final-assembly firms and intermediate-input manufacturers. The second reason is the narrowing of the income gap between the United States and other advanced countries. The domestic markets of European countries and of Japan grew enough to become the first targets for US MNEs. In addition, European and Japanese MNEs grew sufficiently to innovate their own products. Accordingly, it was possible for initial production to start in Europe or Japan and not necessarily in the United States.

Besides these two factors pointed out by Vernon, there may be several other factors affecting the time lag in the transfer of production from parents of the MNEs' to their foreign affiliates. For instance, the accumulation of the MNEs production experiences by the foreign affiliates for a certain product line is likely to reduce various setup costs when production of new items in the same product line is initiated. Further, the change in the factors associated with trade policy, such as import quota and import tariff, by developing countries as well as by the third country can obviously affect the MNE's decision on production location.

As well as looking at these factors which affect the time lag in production transfer, I will in the following analysis focus on factors which are associated with the change in the relative cost conditions between the Japanese parents and their Asian affiliates, so that the impact of the dynamism arising from the remarkable growth performance in the Asia-Pacific Basin on the restructuring behaviour of Japanese MNEs can be identified. Vernon's discussion (1979) on the narrowing of income gap between the United States and other advanced countries will be extended to the recent context of the Asia-Pacific Basin, by applying my more formally developed model (Hayashi, 1992).

3.2.3 *The narrowing gaps in the stages of economic development and the change in the relative cost conditions*

As a result of the recent growth performance in the Asia-Pacific Basin, the gap in the stage of economic development between Japan and other countries in the region seems to have been narrowed. This implies that the difference in cost conditions between the Japanese parents and their Asian affiliates also narrows down, which will affect t^*, or the transfer lag of items. In particular (with reference to the model in Hayashi, 1992, chapter 3), it is possible to specify four production factors: the amount of unskilled labour; the amount of skilled labour; input to obtain communication services; and input to obtain transportation services. Then, the narrowing of the cost conditions of Japanese parents relative to Asian affiliates can be explained as follows.

In the Asia-Pacific region, Japan has gone through a longer experience of industrialisation and has reached a more advanced stage of economic development than the other countries under study. As a result, Japan is currently endowed with a relatively abundant supply of skilled labour and a relatively scarce supply of unskilled labour, whereas other countries in the region are still endowed with a relatively abundant supply of unskilled labour and a relatively scarce supply of skilled labour. Accordingly, it is expected that the ratio of the wage of unskilled labour in Japan (w_J) to that in Asia (w_A), W, is relatively high, and its value is higher than the ratio of the wage of skilled labour in Japan (u_J) to that in Asia (u_A), U.

However, going through a rapid growth and proceeding quickly through the stages of economic development, Asian countries have expanded their labour-intensive manufacturing sectors which has absorbed the abundant supply of unskilled labour. As a result, the wage level w_A has started rising, which tends to decrease the ratio W. However, with more experience in industrial production and employment, these countries have created and accumulated a supply of skilled labour. Accordingly, the wage level u_A is lowered relative to u_J such that the ratio tends to increase.

In terms of market size, measured by per capita and total GNP, the Japanese market has far exceeded those of other countries. In addition, in Japan there are more intermediate-product suppliers with high levels of technological capability that have established close business relations with the final-product manufacturers.

Thus, compared with their Asian affiliates, the Japanese parents of final-product manufacturers are more likely to be assured of geographical proximity to the world market as well as to suppliers of intermediate products. Accordingly, the ratio of communication inputs in Japan (v_J) to those in Asia (v_A), V, and the ratio of transportation inputs in Japan (z_J) to those in Asia (z_A)[9], Z, are relatively low, and their values are lower than the relative wage of unskilled labour, W.

However, as they have proceeded through their stages of economic development, the income level and the size of the market of Asian countries have expanded steadily, even faster than in Japan. Thus, compared with Japan, Asian affiliates are expected to be closer geographically to a greater portion of their markets. Moreover, with the evolution of economic development, more manufacturers of components and other intermediate products will become locally available to Asian affiliates. Accordingly, for Asian affiliates, the unit price for communication inputs (v_A) and for transportation inputs (z_A) are lowered such that V and Z are raised.

By using comparative statics already developed (Hayashi, 1992, chapter 3), one can see that the narrowing of the gap in the stages of economic development and the resulting changes in the relative cost conditions will affect t^* as follows (see also figure 2).

First, the declining trend of the wage ratio W (w_J/w_A) shifts down the $D(t)$ curve so that t^* is lengthened, the rise in wage w_A reducing the cost condition for Asian affiliates at all stages of the product life cycle. However, the rising trend of the ratio $U(u_J/u_A)$, V (v_J/v_A), and Z (z_J/z_A) shift up the $D(t)$ curve so that t^* is shortened, as the decline in u_A, v_A, and z_A improve the relative cost conditions for the Asian affiliates at all stages of the product life cycle. In sum, the total effect of the change in cost conditions on t^* depends on the relative sizes of the decline in W, which lengthens t^*, and the rise in U, V, and Z, which shorten t^*.

3.2.4 *Recent changes in relative cost conditions*

What, then, about the empirical evidence for the total effect of these changes on relative cost conditions? Although precise data are not available, the following information strongly suggests that, although the declining trend in the wage ratio for unskilled workers, W, and the rising trend in the wage ratio for skilled workers, U, are not very definite, the rising trend in the communication cost ratio, V, as well as the transportation cost ratio, Z, seems substantial. In other words, because of the expansion of the domestic market and the development of local intermediate-product manufacturers, the costs of the inputs for communication and transportation services for Asian affiliates have been declining, which seems to have had a significant impact on shortening t^*.

First, as proxies for wages of unskilled labour (w_J and w_A), the average wage payments to employees overall in the electric machinery industry of Japan and the Asian developing countries are examined.

The long-term trend of W since the 1970s seems to have been as predicted in the narrowing-gap discussion; that is, in terms of the domestic currencies, deflated by GDP, the real growth rates of wage payments in Asian NIEs have exceeded those in Japan since the 1970s, and those in the ASEAN-4 countries have exceeded Japan's since the 1980s (table 10).

---

[9] Here, z_i is used to indicate the unit cost of T (transportation inputs) in location i, which is parallel to the costs of other production factors. However, in the original model discussed in Hayashi (1992), the unit cost of T in location i is specified to be a specific ratio, μ_i, of the total production cost, and the index of the relative transportation cost in Japan to that in Asia is defined as $Z = \mu_J/\mu_A$, which is not parallel to those of other production factors, and is mathematically more complicated. The major conclusions of the model remain unchanged in either specification. Some rationale for the specification in the original model is discussed in Hayashi (1992, pages 45-50).

However, the rapid appreciation of the Japanese yen since 1985 has greatly increased *W*, which has given a significant blow to the narrowing-gap discussion. This is shown in table 11, which gives the recent trend in the relative wage, *W*, of Japan to each of the Asian NIEs and ASEAN-4 countries as well as the exchange rate of the Japanese yen over the US dollar.

Table 10. The growth rate[a] of the real wage, in domestic currency (sources: DGBAS, various years; EIAJ, 1989; UN, various years; World Bank, various years).

	1970–80	1980–89		1970–80	1980–89
Japan	4.54	2.58	Malaysia	−4.14	11.10
Korea	6.67	4.76	Indonesia	5.35	na
Taiwan	5.31	7.57	Philippines	−1.56	2.60
Hong Kong	na	0.58	Thailand	na	7.58
Singapore	4.98	5.25			

na not available.
[a] The calculated values are the annual growth rates per annum, deflated by GDP.

Table 11. The exchange rate of the Japanese yen, and the wage of unskilled employees in the electric machinery industry in Japan (w_J) relative to that in Asian countries (w_A) (source: see table 10).

	1970	1975	1980	1985	1986	1987	1988	1989[a]
Exchange rate:								
Yen per US dollar	360.0	296.8	226.7	238.5	168.5	144.6	128.2	138.0
Relative wages (w_J/w_A)								
Korea	3.85	6.44	3.28	3.62	5.37	5.02	4.37	3.38
Taiwan	4.11	4.92	4.38	3.48	4.50	4.08	3.80	2.90
Hong Kong	na	2.76	2.58	2.88	3.71	3.92	5.36	4.59
Singapore	2.66	2.52	3.18	1.91	2.88	3.36	3.74	3.58
Malaysia	2.26	2.23	6.53	3.87	6.05	na	6.55	5.95
Indonesia	13.37	8.65	10.63	9.98	16.03	na	na	na
Philippines	3.16	7.21	7.13	7.59	11.82	11.05	13.81	12.14
Thailand	na	na	14.69	10.14	15.48	na	16.56	15.49

na not available.
[a] Values predicted by EIAJ (1989).

In addition, the recent economic boom in Japan has caused a shortage of unskilled labour, which has further raised *W*. With these mixed effects, the future trend of *W* is difficult to predict.

Second, in terms of skilled labour, some of the interviewees in this study indicated that the accumulation of production and employment experiences by the Asian affiliates has created a significant pool of skilled labour who are well-acquainted with firm-specific knowledge. For instance, sufficient skilled labour has been trained in affiliates of firm P (case 18), where production began more than fifteen years ago. As a result, improvement in production technology such as the modification of capital equipment and rearrangement of the flow of the production process are continuously carried out in the Asian affiliates. Some of these modified technologies are now implemented even by the Japanese parents.

However, the survey results conducted by MITI (1986; 1989b; 1990) indicate that the dramatic expansion in the scale of production and employment in the Asian affiliates of Japanese MNEs in recent years parallels the increase in the demand for skilled labour, where the pace of the increase in demand is even higher than that of the increase in supply. As a result, Asian affiliates have to be more dependent on skilled labour from the Japanese parents, which keeps the wages of skilled labour in the Asian affiliates (u_A) high. However, in the near future, it is expected that much more local skilled labour will become available in the Asian affiliates and thus the wages of such labour will be lowered.

Third, associated with the geographical proximity of the market, final-product manufacturers, and intermediate-product manufacturers, the prices of communication and transportation services (v_A and z_A, respectively) seem to have decreased significantly in the Asian affiliates.

Since 1970, the size of the domestic market for electric machinery as well as for other manufactured products in the Asian NIEs and ASEAN-4 countries have grown at much higher rates than in Japan and the United States.(10) This implies that the Asian affiliates are geographically closer to the market, which reduces their costs for communication and transportation inputs.

In addition, survey results such as in MITI (1986; 1989b; 1990) and Yahata (1990) indicate that the local-content ratio of Asian affiliates for electronic components rose significantly in the 1980s as a greater number of component manufacturers became available in Asia. This implies that the Asian affiliates have achieved greater geographical proximity to input suppliers, which again reduces the costs for communication and transportation inputs.

In sum, the information described above suggests that, although the declining trend in W and the rising trend in U are not clear, the rising trends in V and Z seem to have been substantial. In other words, because of the expansion of the domestic market and the development of local intermediate-product manufacturers, the costs of inputs for communication and transportation services for the Asian affiliates have been declining. Then, by using the result from the comparative statics, I suggest these trends of rising V and Z have a significant impact on shortening t^*.

4 Summary, and issues for further research

Examining the intrafirm division of labour between the Japanese MNE parents and their Asian affiliates in the electric machinery industry, I illustrated in section 2 that the restructuring of items within a product line has been the dominant pattern in their restructuring behaviour. In particular, the two characteristics of this pattern take the form of vertical and horizontal restructuring and were described in detail for the product lines of VCRs and capacitors.

Showing some supporting evidence, I presented in section 3 the third characteristic of this pattern of restructuring behaviour, that is, a decreasing time lag in production transfer from the Japanese parents to their Asian affiliates. Further, using the framework developed earlier (Hayashi, 1992), I explained this trend in association with the change in the relative cost conditions between the Japanese parents and their Asian affiliates, changes arising from the recent high-growth performance in the Asia-Pacific Basin.

(10) Defining the size of the domestic market to be domestic production plus imports minus exports, I have calculated for recent years the sizes of the domestic market for the Asian NIEs and ASEAN-4 countries as well as for those of Japan and the United States (Hayashi, 1992, chapter 6).

To sum up, in the recent dynamism in the Asia-Pacific Basin, the horizontal restructuring of Japanese MNEs seems to have been accelerated in terms of production transfer. In the notation introduced in section 1, it implies that, because of the shortened t^* and the negative value of dt^*, h_A tends to be shorter than h_J. In other words, the product upgrading is carried out by Asian affiliates at an interval, h_A, which is shorter than that experienced by Japanese parents, h_J. Thus, by examining the recent trend in t^*, I have revealed the recent trend in the relationship between h_J and h_A, but the length of h_J has not been explicitly examined.

Can it also be asserted, then, that h_J has tended to be shortened in the recent dynamism in the Asia-Pacific Basin? Or, to put it differently, has the vertical restructuring, or product upgrading, by Japanese parents been accelerated by the acceleration of horizontal restructuring? According to Hara (1990), MITI (1989a), and EPA (1990), this seemed to have been the case during 1987 and 1990. Fueled by active R&D and domestic investment, Japanese parents have accumulated their managerial resources and achieved their product upgrading so that the production and employment in Japan has been maintained, although the recent recession, along with the collapse of the stock market, might reverse this trend.

In any case, a more systematic analysis of vertical and horizontal restructuring behaviour is needed. In particular, questions such as the following should be investigated: how does the coordination of horizontal and vertical restructuring vary across industries, firms, and source countries of MNEs?; and for the Japanese parents, what determines the relative size of horizontal and vertical restructuring, given that some MNEs are more focused on horizontal restructuring, or production transfer to Asian affiliates, whereas others are more focused on vertical restructuring, or product upgrading. These and other questions should be the subject of future empirical and theoretical research.

References

Abernathy W J, Clark K B, Kantrow A M, 1983 *Industrial Renaissance: Creating a Competitive Future for America* (Basic Books, New York)
Asakura R, 1990, "Family high-tech" *Asahi Gendai Yougo* (Asahi Shinbunsha, Tokyo)
Baba Y, 1989, "Characteristics of innovating Japanese firms—reverse product cycles", in *Technology Strategy and the Firm: Management and Public Policy* Ed. M Dodgson (Longman, Harlow, Essex) pp 25–37
DGBAS, various years *Statistical Yearbook of the Republic of China* various issues, Directorate General of Budget, Accounting and Statistics, Li-Ming Cultural Enterprise Co. Ltd, 3, 10th Floor, Hsin-yi Road, Sec. 1, Taipei
Dunning J H, 1979, "Explaining changing patterns of international production: in defence of the eclectic theory" *Oxford Bulletin of Economics and Statistics* **41** 269–295
EIAJ, 1989 *Tounan Ajia Denshi Kougyou Chousadan Houkokusho* (Report on field research of electronic industries in southeast Asian countries) Electronic Industries Association of Japan (Buhin Unei Iinkai, Tokyo)
EIAJ, 1990a *1989 Nen niokeru Denshi Kougyou no Doukou* (Trends in the electronic industry in 1989) Electronic Industries Association of Japan, Tokyo Chamber of Commerce and Industry, Building 3-2-2, Marunouchi, Chiyoda-ku, Tokyo 100
EIAJ, 1990b *Minseiyou Denshi KiKi Data Shu* (Data book on consumer electronic equipment) Electronic Industries Association of Japan, Tokyo Chamber of Commerce and Industry, Building 3-2-2, Marunouchi, Chiyoda-ku, Tokyo 100
EPA, 1990 *Keizai Hakusho* (White paper on the economy) Economic Planning Agency (Japan Government Printing Office, Tokyo)
Finger J M, 1975, "A new view of the product life cycle theory" *Weltwirtschaftliches Archiv* **111**(1) 79–99
Freeman C, 1982 *The Economics of Industrial Innovation* (MIT Press, Cambridge, MA)
Giddy I H, 1978, "The demise of the product cycle model in international business theory" *Columbia Journal of World Business* **13**(1) 90–97

Gruber W, Mehta D, Vernon R, 1967, "The R&D factor in international trade and international investment of United States industries" *Journal of Political Economy* 75(1) 20-37

Hara M, 1989, "Kigyounai sangyou chousei: takakuka to chokusetsu toushi" (Intrafirm industrial adjustment: diversification and foreign direct investment) *World Economic Review* 33(8) 25-31

Hara M, 1990, "Japanese industrial adjustment and foreign direct investment", mimeo, Faculty of Economics, Kobe University, Kobe

Hayashi T, 1992 *Intrafirm Transfer of Production along an Item's Life Cycle: A Case Study of Japanese Multinational Enterprises in Electric Machinery Industry* PhD dissertation, Department of Economics, University of Hawaii at Manoa, Honolulu, HI

Hirsch S, 1967 *Location of Industry and International Competitiveness* (Oxford University Press, London)

Hirsch S, 1972, "The United States electronics industry in international trade", in *The Product Life Cycle and International Trade* Ed. L T Wells Jr (Harvard University Press, Cambridge, MA) pp 39-52

Hirsch S, 1975, "The product life cycle model of international trade: a multi-country cross section analysis" *Oxford Bulletin of Economics and Statistics* 37 305-317

Keesing D B, 1967, "The impact of research and development on United States trade" *Journal of Political Economy* 75(1) 38-48

MITI, 1986 *Kaigai Jigyo Katsudo Kihon Chosa: Kaigai Toshi Tokei Soran* (A comprehensive survey of foreign investment statistics) number 2, Ministry of International Trade and Industry (Keibun Shuppan, Tokyo)

MITI, 1989a *Tsusho Hakusho* (White paper on international trade), Ministry of International Trade and Industry (Japan Government Printing Office, Tokyo)

MITI, 1989b *Kaigai Jigyo Katsudo Kihon Chosa: Kaigai Toshi Tokei Soran* (A comprehensive survey of foreign investment statistics) number 3, Ministry of International Trade and Industry (Keibun Shuppan, Tokyo)

MITI, 1990 *Wagakuni Kigyo no Kaigai Jigyo Katsudo* (Overseas activities of national firms) nos 18-19, Ministry of International Trade and Industry (Japan Government Printing Office, Tokyo)

Shimura Y, 1986 *Denshi Buhin* (Electronic components) (Nihon Keizai Shinbunsha, Tokyo)

Stobaugh R B, 1972, "The neotechnology account of international trade: the case of petrochemicals", in *The Product Life Cycle and International Trade* Ed. L T Wells Jr (Harvard University Press, Cambridge, MA) pp 83-105

UN, various years *Industrial Statistics Yearbook* various issues (United Nations, New York)

Urata S, 1991, "The rapid increase of direct investment abroad and structural change in Japan", in *Direct Foreign Investment in Asia's Developing Economies and Structural Change in the Asia-Pacific Region* Ed. E D Ramstetter (Westview Press, Boulder, CO) pp 175-199

Vernon R, 1966, "International investment and international trade in product cycle" *Quarterly Journal of Economics* 80 190-207

Vernon R, 1979, "The product cycle hypothesis in a new international environment" *Oxford Bulletin of Economics and Statistics* 41(4) 255-267

Walker W B, 1979 *Industrial Innovation and International Trading Performance* (JAI Press, Greenwich, CT)

World Bank, various years *World Tables* various issues (The John Hopkins University Press, Baltimore, MD)

Yahata S, 1990, "Audio-Kiki Sangyo no Kokusai Buhin Choutatsu Taisei" (The international procurement system of intermediate components in the audio devices industry for Japanese firms", in *Kikai Sangyo no Kokusaika to Buhin Chotatsu* (The internationalisation and procurement of intermediate components in the machinery industry) Ed. K Kitamura, Institute for Developing Economies, 42 Ichigaya Hommura-cho, Shinjuku-ku, Tokyo 162, pp 27-64

© 1994 a Pion publication printed in Great Britain

195
199
17
WHARTON REPROGRAPHICS'

Environment and Planning A 1994, volume 26, pages 33-51

LIPPINCOTT LIBRARY
DOCUMENT DELIVERY SERVICE

Industrial restructuring, the Singapore city-state, and the regional division of labour

K C Ho
Department of Sociology, National University of Singapore, 10 Kent Ridge Crescent, Singapore 0511, Republic of Singapore
Received 30 March 1992; in revised form 15 January 1993

Abstract. With continuous economic growth in the 1970s and 1980s and increasing overseas competition, Singapore, as a small city-state with an open economy, faces perhaps the greatest challenge among the Asian newly industrialised economies in their attempts to maintain the pace of development. The effects of a chronic labour-supply situation and the appreciating Singapore dollar on the export competitiveness of the manufacturing sector are reviewed. Enterprise and state responses to these mounting pressures are examined, particularly with regard to the labour shortage. One optimistic long-term solution seems to lie in regional cooperation in industrial development, with a sectoral and technical division across Singapore, Johor in Malaysia, and the Riau Islands of Indonesia. With the three regions in different stages of development, the regional development plan is to bring about a division on the basis of the relative availability of land, labour, and infrastructure. In the second half of the paper, the Singapore–Johor link, the most developed side of the growth triangle is examined, and the potential and problems arising from this arrangement are explored.

1 Introduction

In an early stage of its industrialisation strategy, Singapore promoted the island as a platform for exports for multinational corporations (MNCs). Tariff protections were dismantled a year after Singapore's independence, when it was realised that the separation of Singapore from Malaysia in 1965 effectively sealed off Malaysia as a potential home market for Singapore's domestic industries.

The entry of MNCs into Singapore in the late 1960s and into the 1970s can be explained by external and internal factors. The entry of foreign capital into Singapore was part of a general movement of First World investments overseas in search of low-cost production sites. This movement was caused by a number of factors—what is termed in the new-international-division-of-labour literature as the declining conditions of valorisation (Cohen, 1981; Frobel et al, 1980; Henderson, 1986)—oversupply of manufactured goods, rising real wages, and the erosion of international regulatory arrangements which supported postwar production and trade (Rich and Linge, 1991).

A number of domestic factors were also responsible for making Singapore an attractive production site. After the transition from colonialism, the internal political situation, that had been marred by student riots, union action, and manoeuvering for dominance by various political parties, began to stabilise by the late 1960s. This period also coincided with the establishment of a set of government-related regulatory institutions to facilitate economic development. At the same time, there occurred a gradual upgrading of the infrastructure, which was already well developed as a result of Singapore's legacy as a colonial port city. With the internal environment conducive for industrial production, there was a rapid growth of foreign direct investment (FDI) into Singapore. In 1975, for example, FDI in Singapore accounted for 48.6% of all FDI in Asia (Mirza, 1986).

This investment wave was led by the United States, which accounted for about half of the foreign investments in Singapore, with Japan contributing about 15%, and the European Community about 23%. It is important to note that foreign dominated over local investments. For the period 1970–80, foreign investments in Singapore accounted for between 70% and 80% of total investments. This created a situation where foreign capital employed a major share of the industrial work force and shaped much of the industrial landscape on the island.

The size of foreign investments, the speed of development in the 1970s, together with the small size of Singapore, produced a highly volatile situation, increasing the vulnerability of the local economy to external shocks because of the export orientation, and at the same time making the economy highly susceptible to internal constraints of land and labour.

The vulnerability of Singapore, to the global economy in general and to the economies of its major investors in particular, can be deduced from the following comment made in October 1988 by the Minister of Trade and Industry:

> "The US watches the US composite index of leading indicators anxiously, to find out whether a downturn is coming. We also watch the same index anxiously, because we have found out that the US index predicts two thirds of Singapore's economic performance, that is assuming we do the right things. The lag between the index and our performance is 2 to 3 quarters. The lag between the index and the US performance is also 2 to 3 quarters. So, we are contemporaneous" (Lee H L, page 3).

In this paper, I will analyse the problems of having to adjust to external and internal forces in order to maintain export competitiveness, identifying the problems which are acute in the context of the Singapore city-state. The rapid growth of FDI in Singapore in the 1970s led to a dramatic change in the economic scenario. In the 1960s and 1970s, by being one of the first developing countries with an explicit export-oriented industrialisation strategy, Singapore was able to promote economic growth simply by selling the island as a low-cost production site. As was pointed out in the Strategic Economic Plan, in the 1980s the field of competitors in this niche broadened considerably (MOTI, 1991a). Moreover the tight situation over land and labour has created two other important considerations for policymakers. Constraints on domestic labour supply in particular have meant that the promotion of industries has become something of a trade-off; the growth of some industries can often mean the demise of others. Also, the need to expand production into Singapore's neighbours (namely Johor, the southern state of Malaysia, and Riau, a cluster of Indonesian islands close to Singapore) in an attempt to overcome land–labour shortages means that further attempts to sustain economic growth have to develop in parallel with efforts to develop and enhance regional relations and administration.

2 Triggers of industrial restructuring in the 1980s

2.1 Labour cost, productivity, and international competitiveness

The flow of FDI into Singapore, a trickle that turned into a flood during the 1970s, quickly soaked up the surplus labour that had built up in the 1960s. The unemployment rate fell steadily from 6% in 1970 to under 4% in 1977. By the late 1970s, Singapore had achieved full employment. With this came the problem of how to ensure an adequate supply of labour to maintain the industrialisation process. Schemes to increase the supply of domestic labour, however, have their limitations.

Part-time employment in 1989 accounted for only 2.6% of the labour force, marginally higher than the 2.1% in 1981 (*Straits Times* 1989e). Although there is potential in increasing this supply, particularly by attracting housewives, the present number remains a small proportion of the labour force.

As for increasing the retirement age of workers from 55 years to 60 years, the government has presently opted not to make the limit of 60 years compulsory, preferring to work through the unions. Although there has been some success, the process is hampered by a seniority wage system which discourages companies from retaining expensive older workers. The government, recognising this feature, has acted recently to reduce the wage costs of workers aged above 55 years by lowering the rate of contribution to the Central Provident Fund (CPF, the compulsory savings fund for workers and employers) (*Straits Times* 1988b). The participation rate of 60–64 year olds has remained stable at around 31%, but the participation rate for those aged 65 years and older has dropped throughout the 1980s, only stabilising in 1990, perhaps because of the lowering of the CPF rates (DoS, 1990).

Of the attempts to increase the domestic labour force, the one targeted at females seems to have worked well. The female labour-force participation rate increased from 29.5% in 1970 to 44.3% in 1980, growing to 48.4% in 1990, but this has been offset by a slight decline in the male participation rate. As a result, the overall labour-force participation between 1980 and 1990 hovered at around 63% (MOTI, 1991b).

Faced with a slow-growing domestic labour force, firms have expressed a greater willingness to supplement their work force by using foreign labour rather than part-time labour or older workers. Managers argue that foreign workers are less likely to job hop because their work permit limits their mobility. In one survey covering seventeen out of eighteen firms operating in Bedok New Town, a public housing estate in southeastern Singapore, the seventeen companies recruited 7897 workers over a ten-month period and lost 7060 workers in the same period. This works out to a loss of seven of every eight workers hired. Because of work-permit restrictions, the turnover rate among foreign workers is much lower. The companies hired 4815 foreign workers and lost only 2811 in the same period (*Business Times* 1989a). Moreover, firms claim that foreign workers are preferred because of their positive work attitudes (including willingness to work extra hours and night shifts), higher education levels, and relative youth. Electronics companies, for instance, find it difficult to use older workers to do scope work, the assembly of miniature components which is an essential part of the production process (*Straits Times* 1989a).

With a slowdown in the increase in local workers, foreign workers moved in rapidly to fill in the gaps. The foreign labour force constituted one third of the growth in the work force between 1975 and 1979. In 1980, out of a work force of 1.07 million, 7.4% (79275) were foreign workers on employment passes and work permits (DoS, 1981 estimated from table 12). Between 1980 and 1984, more than half of the work-force increases were accounted for by foreign workers (MOTI, 1986a).

Alarmed by the rapid growth in number of foreign workers in the late 1970s and early 1980s, the government considered phasing out the dependence on unskilled foreign workers by 1990 (Pang and Lim, 1982). The worry was that the presence of a large population of foreign workers might cause social and political problems (MOTI, 1986a). This intention to phase out foreign workers has not, however, been realised, for a number of reasons.

(1) Companies, for various reasons mentioned earlier, continue to prefer foreign labour over older and part-time labour. These users argue that foreign labour has

kept costs down. With the labour shortage, even the costs of recruitment have increased. Seagate Technology, one of the largest employers in Singapore, is estimated to have spent in 1987 S$400000 in one local newspaper alone on advertisements for workers (*Straits Times* 1988a). Over the years, through the various chambers of commerce, the Manufacturers Association, and the Employers Federation, employers have not only supported the use of foreign labour but also voiced their opposition towards paying for increases in levies on foreign workers.

(2) Supplementary labour streams (part-timers, older workers) are not flexible enough to expand or contract in tandem with production orders. Such labour cannot be increased substantially within a short period. Increasing the part-time employment pool mainly involves persuading housewives to work, and this requires a long-term effort to change attitudes, as well as to balance the economic and social costs of providing child care outside the family. Aside from increasing the costs for employers, having a greater number of older workers also involves changing the attitudes of older workers as they must be persuaded to delay retirement plans made attractive by the sizable savings resulting from the compulsory savings scheme and to accept the lower status and lower paying jobs in reemployment and continuing employment.

(3) There is also the question of motivation and the relative productivity levels of the different types of labour. Older workers are less likely to take the unpopular night shifts. A Ministry of Labour (MoL, 1989) survey of males aged 50 years and over indicated that the availability of simple and/or light work and of short flexible working hours were the most important considerations for older people to continue working or to start working again.

Given these considerations, the government's original position in the late 1970s and early 1980s of viewing foreign workers as a temporary stopgap measure has changed to one which allows for a carefully controlled intake. In 1989, although

Table 1. Deployment of foreign workers in US companies in Singapore, 1990 (source: US Embassy, Singapore, reproduced in *Business Times* 1990d).

Sector	Number of respondents	Source of workers (%)			
		US	Singapore	other	total
Manufacturing:					
electronics	61	0.3	74.1	25.6	50881
others	56	0.2	76.8	23.0	11224
Chemicals and pharmaceuticals	48	0.2	79.5	20.3	3049
Aerospace and aviation	12	1.2	91.6	7.2	2368
Finance	60	1.8	92.3	6.1	4173
Consultancy and engineering services	33	3.2	65.5	31.3	2175
Petroleum	5	1.8	83.4	14.8	1954
Oil-related industries:					
equipment manufacturing	14	3.2	81.6	15.2	1084
sales and services	63	5.5	80.4	14.1	1555
Travel, transportation, tourism, and hotels	21	1.1	98.2	0.7	1327
Advertising and public relations	14	2.4	90.4	7.2	595
Other services	77	2.1	94.2	3.7	8796
Sales (including representative offices)	149	3.9	92.3	3.8	4440
Legal services	4	11.8	68.5	19.7	76
Total	617	1.0	79.2	19.8	93717

Note: effective statistical base, 792 companies; effective response rate, 77.9%.

continuing to caution against an overreliance on foreign workers, the Labour Minister mentioned, as government policy, that foreign workers should be used as a buffer to even out swings in the business cycle (*Straits Times* 1989b). Most significantly, the use of foreign workers as a strategy to overcome labour imbalances was advocated in the 1991 Strategic Economic Plan (MOTI, 1991a).

Accordingly, the government has developed foreign-worker policies that are more indirect, using the foreign-worker levy and quotas to constrain the growth in number of foreign workers. The sizes of the levy and quotas differ for skilled and unskilled workers, and also vary for different industries.

In mid-1989, the government also attempted a longer term solution to boost the labour supply by relaxing restrictions for permanent residency in Singapore. Originally targeted at technicians and blue-collar workers in Hong Kong, the new standards allow those with a secondary school and technical and/or vocational education to qualify and gives successful applicants the option to rent public housing as well as a five-year period to make up their minds. The Singapore High Commissioner to Hong Kong was quoted as saying that over the next five to eight years Singapore will allow in as many as 100 000 Hong Kong residents (*Straits Times* 1989c).

A 1989 estimate placed the existing number of foreign workers at about 150 000 (*Straits Times* 1989d). Without the benefit of current official statistics on foreign workers, the most recent comprehensive picture of the deployment of foreign workers is derived from a survey of US companies released in 1990 by the US embassy in Singapore (see table 1). The employment pattern among US companies indicates that foreign workers, including executive, skilled, and unskilled workers, are concentrated in manufacturing and certain kinds of services (consultancy and engineering services).

The constraints to enlarging the labour supply inevitably resulted in a wave of wage increases. These pressures for wage increases in the manufacturing sector are being worsened by the volatile external demand faced by labour-intensive, export-oriented industries such as the electronics and apparel industries. As shown in table 2, the rapid employment growth in the electronics industry in the 1980s suffered two shocks, one in 1982 when the labour force was trimmed to about 60 000 workers, and another in 1986 when the electronics work force dropped from

Table 2. Key indicators in the electronics and garment industries in Singapore, 1980–88 (sources: DoS, 1980–88).

	1980	1982	1984	1986	1988
Electronic products					
Number of workers	71 506	60 140	72 380	69 332	112 769
Percentage of total manufacturing work force	24.9	21.7	26.3	28.1	34.7
Direct exports as a percentage of output	84.6	86.3	85.0	88.7	86.0
Number of establishments	170	182	209	187	239
Garments					
Number of workers	27 188	28 608	25 710	24 812	29 402
Percentage of total manufacturing work force	9.5	10.3	9.4	10.0	9.1
Direct exports as a percentage of output	71	70.2	73.6	80.1	83.9
Number of establishments	374	406	387	370	370

72380 in 1984 to 69332 in 1986. The rebound after 1986, spurred by the growth of the computer industry, led to an addition of over 40000 workers within a two-year period.

Similarly, though on a smaller scale, the garment industry, which is the second largest employer in the manufacturing sector, experienced a downturn in the period 1985–86, resulting in a loss of about 3000 workers from the industry. This increased to 29402 workers towards the end of the 1980s.

The increasing dominance of the electronics industry in Singapore in terms of employment and the value of direct exports implies not only that the manufacturing work force will continue to be buffeted by variable external demand but also, more importantly, that such shocks are likely to be felt by an increasing proportion of the work force.

With a tight labour market, a rapid increase in external demand such as the one experienced by the electronics industry in the period 1983–84 not only creates a pressure to increase wages but also prevents companies from maintaining the skill levels of production operators. This is especially serious in Singapore because of its drive to upgrade the technology and skill levels.

The technological upgrading in the 1980s has arisen in part from the entry of new industries, such as computers, chemicals, and precision engineering (Natarajan and Tan, 1992) as well as from the upgrading of the manufacturing process through automation and introduction of higher-value-added lines (Mirza, 1986; Rodan, 1989). The state-initiated drive to increase technical skills from vocational to tertiary levels has resulted in an increase in the education levels of the work force; the percentage of workers with postsecondary education increased from 12.5% in 1980 to 17.0% in 1986 (MOTI, 1986b). Many of these workers are likely to possess technical diplomas and trade certificates, a reflection of the increased output of the polytechnics and vocational institutes.

The result is a production process that is intermediate in terms of technological sophistication and skill level. The combination of these characteristics, coupled with a volatile export market and a tight labour situation, mean that with any sudden increase in labour demand, unless the overtime option is taken, employers may be forced to rely on supplementary labour streams (usually older housewives who work part time) or to poach workers from other industries (usually from the garment industry). Either way, productivity levels are likely to decrease because of the slower work pace of older workers and the adjustment problems of workers not used to production lines (*Business Times* 1989b).

These reasons mean that in a tight labour-market situation, any sustained increases in labour demand not only increases wages but also lowers productivity.

Table 3. Wages, total labour costs, and productivity changes in Singapore, 1973–84 (source: MOTI, 1986a).

Period	Average monthly earnings		Total labour costs		Real productivity growth
	nominal	real[a]	nominal	real[b]	
1973–78	9.7	1.7	9.1	1.2	2.7
1979–81	11.9	4.7	14.4	7.4	4.8
1982–84	11.2	8.3	12.5	12.8	4.3

[a] Consumer Price Index used as deflator.

[b] A deflator that measures the product price of each sector is used.

The influential Economic Committee, which was formed to review the progress of the Singapore economy and to identify new areas for its future growth, identified the increasing wage costs, unmatched by productivity growth, as one of the main causes of the 1985 recession in Singapore (see MOTI, 1986a). Table 3 indicates that the adjusted earnings and labour-cost rate lagged behind the rate of productivity growth over the period 1973–78. However, with the tightening labour market, wage increases matched productivity gains in the period 1979–81 and exceeded productivity gains in the period 1982–84, precipitating the recession of 1985.

2.2 Automation

The push towards greater automation in the factories was seen as a way of escaping the problem of overdependence on foreign labour. Targeted particularly at smaller firms and those relying on foreign labour, the automation effort is supported by a number of schemes. One important scheme, the Robot Leasing and Consultancy (RLC), was set up in 1984 to allow companies to apply for low-interest loans to finance automation. By the end of 1988, some S$110 million had been given out. This scheme is supported by a consultancy service to companies to help them automate their production. Under the consultancy scheme, the company pays 10% of the costs, with the Skills Development Fund subsidising the rest. Automation equipment can also be leased from the RLC for periods of up to seven years. In 1988, the automation project was further supported by a S$60 million fund administered by the Economic Development Board (EDB). From this fund, S$50 million is to be used for financing the leasing and purchases of automation equipment, and the remaining S$10 million is for grants of up to 70% of the cost of training automation engineers and technicians. Automation projects are also boosted by investment incentives in terms of tax deductions (150% write-off against taxable income) and accelerated depreciation on robotic and CAD/CAM(1) equipment (*Straits Times* 1988c).

A 1989/90 survey commissioned by the Singapore Industrial Automation Association (SIAA), covering 995 companies from a broad spectrum of industries, revealed that with the exception of standard production machines (for example, power presses, machines tools, injection moulding machines) the use of other types of automated equipment (computer numerically controlled machinery, industrial robots, programmable process-control equipment) remained very low. It was found that between 70% and 98% of the companies surveyed had no intention of obtaining these latter types of automated equipment (SIAA, 1990). As shown in table 4,

Table 4. Factors hindering automation in Singapore (sources: SIAA, 1988; 1990).

Factor[a]	Survey	
	1987/1988	1989/1990
Heavy investment and uncertainties	59	48.1
Inadequate in-house experience	47	39.4
Problem of integrating the new and the old system	36	28.0
Lack of confidence in automation-equipment suppliers	27	21.3
High initial operating costs	52	15.8
Lack of financial backing	32	12.9
Lack of training opportunities	27	8.9
Problem of displaced workers	15	4.8
Number of companies responding	222	729

[a] For each factor, the percentage of companies for which that factor was relevant is given.

---

(1) Computer-aided design/computer-aided manufacturing.

some of the reasons companies gave for refraining from undertaking automation projects were consistent with those given in the surveys commissioned in 1987/88 and 1989/90. The most important reason for not undertaking automation was the high cost of investment and the uncertain returns for such expenditures. This reason seems well founded. One conservative estimate indicated that at S$100 000 per robot installation and a saving of one worker at S$1000 per month working on two shifts, the payback period was slightly over four years (*Straits Times* 1987). Many robotic installations cost more than S $100 000.

Thus, although the automation push is designed to solve the problem of the tight labour market, it seems that automation effort and the labour situation in Singapore are linked in contradictory ways. On the one hand, although the level of wages in Singapore is rising, paradoxically it is not high enough to shorten the period of return on automation investments to make companies willing to invest. On the other hand, the lack of trained workers in Singapore hinders automation efforts.

2.3 The appreciating Singapore dollar

With its successful incorporation into the global economy as a result of the entry of MNCs into the country, Singapore began to enjoy consistent balance-of-payments surpluses. As a result, the official reserves grew from S$3.1 billion in 1970 to S$13.7 billion in 1980 (DoS, 1983). The 1980s saw an even faster rate of increase. By 1984, the reserves grew to S$22.7 billion (DoS, 1990). This translates into an appreciating Singapore dollar. Since the Singapore dollar was floated on 20 June 1973, it has been appreciating steadily against the US and Australian dollars. In relation to the currencies of the Asian newly industrialised economies (NIEs), the Singapore dollar has shown a steady appreciation against the Hong Kong dollar and the Korean won, while gaining against the Taiwan dollar only up to the mid-1980s. These changes can be seen in table 5.

The Monetary Authority of Singapore has consistently maintained a strong and stable policy on the Singapore dollar, compatible with the government's desire to build the reputation of Singapore as a regional financial centre (BERI, 1985). As Singapore imports practically all of its goods, the import content for goods manufactured in Singapore is high. Sandilands (quoted by S Y Lee, 1986, page 155) estimated the import content of domestic exports to be as high as 85%. Thus the strong dollar means cheaper imports of raw and intermediate goods, which offsets the higher export prices. On the consumption side, the high exchange rate means lower import costs, which keeps the cost of living down, thereby reducing pressures for higher wage claims (Lee S Y, 1984).

A strong dollar, however, has had different impacts on domestic industries. The EDB quarterly surveys on business expectations, for example, revealed that, although more industries were beginning to feel the impact of the strong dollar in

Table 5. Exchange rates for the Singapore dollar, 1970–90 (average rates for stated years) (sources: DoS, 1980; 1990).

	1970	1973	1976	1980	1983	1986	1990
Singapore dollars per:							
US dollar	3.09	2.44	2.47	2.14	2.11	2.18	1.81
Australian dollar	3.45	3.47	3.03	2.44	1.91	1.46	1.41
Hong Kong dollar	0.51	0.47	0.50	0.43	0.29	0.28	0.23
100 Korean won			0.51	0.35	0.27	0.25	0.26
100 New Taiwan dollar			6.21	5.95	5.27	5.76	6.74

1990 compared with 1988 or 1986, the appreciating dollar was more likely to affect the export orders of the textiles, metal products and machinery, and paper industries. For example, 27.7% of the textile companies and 25.8% of companies in the metal products and machinery industries felt that the strong Singapore dollar was the most important factor limiting their export orders in 1990. There have been suggestions (for example, Bryant, 1989) that the appreciation of the currency contributed in an important way to the 1985–86 recession.

The Economic Committee's (MOTI, 1986a) own position about the role played by the strong dollar in the 1985 recession is less certain. Although the negative effects of an appreciating dollar were noted, the Committee recommended the use of more direct measures for dealing with the problem, rather than through the exchange rate:

> "The Committee is of the view that the Singapore dollar exchange rate should be set by market forces, but the impact of the exchange rate on our export competitiveness and tourist costs should be taken into account ... Rather than attacking the problem through indirect means of changes in the exchange rate, the Committee feels that it is better to confront this difficulty squarely by reducing CPF contributions and holding wage costs down. The conclusion that Singaporeans have to accept a lower standard of living for a time is best put across directly, rather than through the disguise of higher inflation brought about by a lower exchange rate" (page 54).

This position of using the exchange rate to control inflation has been maintained throughout the 1980s. The Finance Minister has recently reiterated a similar position that Singapore's annual inflation rate of 2.8% (1975–91) was one of the lowest in the world, and had the exchange rate not appreciated since 1988 the inflation rate would have been twice as high in 1991 (*Straits Times* 1992b).

The impact of the appreciating Singapore dollar compared with that of labour problems (in terms of availability as well as cost) on business as perceived by manufacturers in Singapore can be appreciated by an examination of table 6.

Table 6. The relative influence of labour and a strong Singapore dollar on production and exports (sources: EDB, various years).

Industry	Labour[a]			Strong Singapore dollar[b]		
	1986	1988	1990	1986	1988	1990
Food, beverage, and tobacco	1.0	4.2	10.2	2.7	6.0	16.7
Textiles, footwear, and wearing apparel	73.7	87.0	94.5	13.6	15.7	27.7
Paper, paper products, printing, and publishing	8.8	15.0	35.9	34.1	4.1	17.3
Building materials	9.5	12.9	43.9	11.2	7.5	15.6
Petroleum refining	0.0	0.0	0.0	0.0	0.0	0.0
Chemical products	0.8	1.5	37.9	3.2	2.6	4.6
Metal and nonmetallic products, and nonelectrical machinery	10.3	35.0	58.9	5.8	2.6	25.8
Electrical, electronic machinery, and appliances	5.8	19.6	23.7	6.0	3.4	7.4
Transport equipment	3.9	21.5	58.4	1.4	0.0	6.8
Other	36.3	54.2	64.8	8.1	2.4	16.2

[a] The percentage of firms agreeing that labour is the most important factor limiting production in the previous quarter (figures represent an average over four quarters).

[b] The percentage of firms agreeing that the strong Singapore dollar is the most important factor limiting production in the current quarter (figures represent an average over four quarters).

It is clear that, with the exception of the high-value-added petroleum industry, an increasing number of manufacturers across the broad spectrum of industries were worried about rising labour costs and were cutting back on production because of it. This sentiment is especially strong for the period 1988–90. The most severely affected was the textile and garment industry, where as any as 94.5% of the firms interviewed identified labour (cost as well as availability) as the most important factor limiting their production.

Compared with labour problems, perceptions about the strong Singapore dollar having an effect on exports seem more mixed. Concern about the strong dollar started surfacing in 1990, shared among a number of manufacturers in the textiles, metal products and machinery, food, and paper industries, but those in the petroleum, chemicals, transport equipment, and electronics industries seemed less affected by the exchange rate.

2.4 Growing regional interdependence: labour movements, firm relocation, and the growth-triangle policy

The impact of Singapore's industrialisation on neighbouring regions was felt most through the influx of Malaysian workers, drawn by the high wages offered by Singaporean factories. This started in the 1970s, when the surplus domestic labour was beginning to dry up as a result of the fast pace of industrialisation. Recruitment agents began to search for female labour from the Malaysian rural sector, and by their efforts created a peripheral work force by pulling in largely unskilled and semiskilled Malaysian workers to supplement the domestic labour supply (Heyzer, 1982). At the peak of the economic boom in 1973, noncitizen work-permit holders numbered over 100000, or about an eighth of the total work force in Singapore (Pang and Lim, 1982). Most of these workers came from Johor, the southern state of Peninsular Malaysia. In spite of tighter work-permit controls and increases in foreign-worker levels, the flow of Johor workers into Singapore continues to be significant. In 1991, the State Committee in Charge of Manpower in the State of Johor estimated that, in addition to Malaysian workers residing in Singapore, there are as many as 35000 workers who commute from Johor daily to work in Singapore (*Straits Times* 1991f). There are also temporary-residence schemes whereby Singaporean companies bus in workers from Johor, to work on twelve-hour shifts, four days at a stretch (Kamil et al, 1991).

In the 1980s, a reverse movement from Singapore to Johor began as companies in Singapore started exploring Johor as an alternative production site to escape the rising costs of labour and land. Boosted by this movement, Johor experienced rapid industrial growth in the late 1980s. In 1988, the Malaysian Industrial Development Authority (MIDA) approved 134 Singaporean projects with a capital input totalling M$172 million.⁽²⁾ Of the total, 78 projects were located in Johor with an equity of M$102 million (*Business Times* 1989c). This investment increased to M$131.5 million in 1989, and to M$370.8 million in the first ten months of 1990. This placed Johor in a competitive position with the state of Selangor, the industrial heartland of Peninsular Malaysia. In 1990, with the number of industrial projects approved by MIDA growing at a faster rate than in Selangor (183 versus Selangor's 210 in 1988, and 208 versus Selangor's 210 in 1989), Johor overtook Selangor to become the top seller of industrial land in Malaysia (*Straits Times* 1990a).

Singaporean investments in Johor tend to be for smaller projects in the rubber, electronics, and textiles industries (*Business Times* 1991a). The small scale is

---

⁽²⁾ M$ is used to denote the ringgit.

evident from table 7, which shows that, although the number of Singaporean projects in Johor far exceeds the number of Japanese projects, Japanese projects (as well as Taiwanese and US projects) tend, on average, to involve more capital.

Table 7. Foreign investment in Johor, Malaysia, January 1981 to October 1990 (source: Malaysian Industrial Development Authority, reproduced in *Business Times* 1991a).

Country	Number of projects	Equity[a]
Japan	119	9217
Singapore	466	7459
Taiwan	107	4495
USA	35	3229
Hong Kong	53	1491

[a] In billions of ringgit.

This evidence seems to indicate that small and medium-sized forms are moving into Johor, a trend encouraged by the Singapore government. A regional location policy seemed to have emerged in the mid-1980s, as the Economic Committee (MOTI, 1986a), pondering the causes of the 1985 recession and the new directions that Singapore should take, argued for the recognition of Singapore's comparative advantage based on its strategic location and its well-developed infrastructure.

This notion of comparative advantage was once commonly used by the colonial administration to describe the advantages of the port city of Singapore and became unfashionable when Singapore started on its industrialisation effort in the 1960s. The resurgence of this idea in the mid-1980s seems to have coincided with the recognition, in the light of constraints (particularly of labour) faced by the small city-state, that the pace of economic growth could not be sustained by industrialisation alone.

Several years later, Goh Chok Tong, then the deputy prime minister, mooted the idea of the triangle of growth to Malaysia and Indonesia, Singapore's immediate neighbours. The growth-triangle concept is based on a technical, sectoral, and regional division of labour. Technically, labour and land constraints require the low-value-added labour-intensive phases of the production process to be subcontracted or relocated to Johor, Malaysia, or Riau, Indonesia, to free these scarce resources for higher-value-added manufacturing activities. In terms of the division of labour by sector, it is also recognised that Singapore plays a more important role in services as a regional headquarters location for MNCs overseeing their plants in Johor or Riau. Such companies would utilise Singapore management, transport, and communication networks and its financial institutions to support their activities. Last, the regional division of labour is clear from the plan, namely that Singapore, Riau, and Johor, as neighbouring regions containing different comparative advantages, can, by playing different economic roles, offer the potential investor a varied but integrated package of potential benefits. The implication that Singapore will end up playing the dominant role in managing this arrangement is not lost on the Malaysians (Kamil et al, 1991) or the Indonesians (Pangestu, 1991). Thus, what is basically an economic arrangement has to be carefully nurtured at the political level. In the next section, I indicate some areas of tension and some benefits arising from regional cooperative efforts. In order to provide a more in-depth analysis, the most developed link of the triangle, the Singapore – Johor link, is chosen for analysis.

3 Regional impacts of industrial restructuring: interdependence versus competition

"Whatever the foreign policies of the two countries [Malaysia, Singapore], Johor, being the closest to Singapore, would be at the centre. As a result, Johor had to be more realistic in dealing with Singapore. I believe, after long years of co-existence, we tend to understand each other better than the people in Selangor" [Tan Sri Muhyiddin, Mentri Besar (Chief Minister) of Johor, in a talk organised by the Singapore Press Club (*Straits Times* 1990b, page 1)].

"You cannot say that Guangdong is not part of China. But we are closer to Hong Kong and Taiwan that we are to Beijing" (ranking executive of a top Guangzhou company, interviewed by *Newsweek* 1992, page 12).

The most significant consequence of industrial restructuring in the late 1980s and into the 1990s for Singapore is clearly in terms of regional effects. The government, in recognising the tendency of firms to relocate production away from Singapore, is promoting a regional restructuring policy in the hope of freeing scarce resources and redistributing these to desired economic activities in Singapore. Such a policy deserves greater attention because of its assumptions on regional relations and cooperation, as well as on location of economic activities and the accompanying division of labour.

3.1 Areas of cooperation

The close ties between Singpaore and Johor can be see from the 1927 water agreement between Sultan Ibrahim of Johor and the Singapore City Council, allowing the city to draw water free of charge from Sungai Pulai. The second water pact, signed in 1961, allowed Singapore to obtain water at 3 cents for every 10000 gallons of water piped into the island, plus rental for land reserved for waterworks. In return, Singapore supplied Johor with purified and/or treated water. A more recent agreement, signed in 1988, allowed Singapore to develop more water resources in Johor at Singapore's expense (*Business Times* 1991d). The arrangement was one of mutual benefit. Singapore, with limited water resources, needed increasing amounts of water, especially with the industrialisation effort. Johor's revenue from the supply of water was M$85 million in 1991, making the sale of water the second largest source of revenue for Johor (*Straits Times* 1991a; 1991b). With a strong historical tie, strengthened further by the water agreement, Singapore–Johor relations are envisaged to become even closer with the development of the growth-triangle concept.

In response to Singapore's plan, Johor's Chief Minister, Tan Sri Muhyiddin Yassin, suggested three areas for cooperation: (1) joint promotions to attract investors, (2) joint development of human resources and the work force, (3) new spheres of cooperation in joint ventures (*Business Times* 1990c). These suggestions led to Singapore's offer to help Johor set up work-force training facilities (*Straits Times* 1991g).

Another significant area of cooperation includes equity stakes in the development of an industrial park in Skudai, Johor. On the Singapore side, Jurong Town Corporation, a government statutory board responsible for developing the first industrial park in the 1960s, will lend its expertise in building and managing the park (*Straits Times* 1990c; 1990e). More recently, Singapore's DBS Land, a subsidiary of the government-controlled Development Bank of Singapore, holding a 30% stake, joined Pernas to build a township of offices, factories, homes, and hotels on a massive site of 1333 ha, 24 km from Johor Bahru, the capital of Johor, and 3 km from Pasir Gudang, the largest industrial estate in Johor (*Business Times* 1991b).

Such cooperative ventures provide further evidence of the diverse roles the state plays in initiating and sustaining economic development among the Asian NIEs (for example, Deyo, 1987; Park, 1991; Rodan, 1989; Wade, 1990). The role of the state in the case of the growth-triangle project is an interesting one. The states involved clearly go beyond their roles of managing their respective internal environments to accommodate capital; these states actively create the conditions conducive for multinational regional cooperation. Such efforts are the consequence of careful diplomacy, as the agreement was forged against a recent past, that involved a confrontation of Indonesia with Singapore and Malaysia between 1963 and 1965, and the separation of Singapore from Malaysia in 1965 under less than amicable circumstances (Turnbull, 1977). This project was also created without the benefit of common cultural ties [a factor that facilitated Hong Kong's participation in Guangdong (Lee T Y, 1992)] and in the midst of continual fears (see next section). By working through their various subsidiaries, states demonstrate further cooperation by providing the capital and management necessary to build the infrastructure and institutions needed to support industrial activities.

3.2 Johor state and Malaysian federal government relations

The linking of Singapore, Johor, and Riau in economic development has led to accusations that the neighbouring regions have become the industrial backyards of Singapore. These statements have been made both by Indonesians [for example, by the Indonesian economist Mari Pangestu (*Straits Times* 1991j)] and by Malaysians [notably by Datuk Shahrir Abdul Samad, a Member of the Malaysian parliament for Johor Bahru (*Straits Times* 1990d)], but the comments are particularly strong on the Johor side. This is because of two important differences between Johor and Riau. Johor has a strong political elite represented in the Chief Minister, who is very much at the forefront in promoting Johor's industrial development. Developments in Riau, however, are guided directly from Jakarta, so much so that one concern is that the provincial government will be left out of the overall planning altogether (Pangestu, 1991). Economically, the Riau islands are very much at the periphery of the Indonesian economy, whereas the state of Johor makes up a large portion of Peninsular Malaysia in terms of population and share of economic activity (Kamil et al, 1991).

These differences mean that, although the Johor government has a direct role in determining its own development, such actions are likely to lead to close scrutiny and criticism from the Malaysian federal government. The water agreement between Johor and Singapore was the source of recent tensions between the state and the federal government when the 1991 water crisis in neighbouring Malacca led the federal government to propose measures for the federal control of water resources. The prospect of ceding the state's water catchment areas, which is estimated to be one third of the Johor territory (*Straits Times* 1991d), and the potential loss of one of the state's largest revenue sources, led to Johor's firm stand against the federal government's proposal for a national water authority.

In the more general area of industrial development, the Malaysian federal government's position has been to guide development in such a way that regional economic imbalances are minimised (Kamil et al, 1991; Perry, 1991). A potential source of federal-state tension is therefore the disagreement over the location of projects. One recent example is the much-publicised, M$8-billion-dollar joint-venture between Taiwan and Malaysia to build the largest integrated steel mill in Southeast Asia. The federal government wanted to locate the mill in Trengannu, in Northeast Peninsular Malaysia, whereas the joint-venture partners wanted it in

Johor because of its better infrastructural facilities and its proximity to Singapore (*Business Times* 1990a; *Straits Times* 1991h). The fact that it ended up in Johor is an indication of the power of capital to dictate the outcomes of such disputes.

Although there may be potential areas of disagreement between state and federal interests, as seen in the water issue and the location of industrial projects, the Johor state will continue to work closely with the federal government in economic development, unlike the case of Guangdong in China (*Newsweek* 1992). This will be so as long as state and federal political elites form a cohesive body. Johor will also conform to the federal government's wishes as long as it is dependent on the federal government for the financing of various infrastructural projects needed to support the state's industrialisation effort and as long as this support is forthcoming. However, as the Guangdong case illustrated, the dependence of the state on federal funding of infrastructural developments is not an inherent problem that cannot be overcome, as with Johor's rapid industrialisation, the state can become rich enough to self-finance infrastructural projects.

Also, although Johor has expressed keen interest in the regional cooperation effort as a way of boosting its industrialisation efforts, some initiatives will have to be developed at the federal government level. The growth-triangle concept, for example, involves the issues of standardising immigration controls and the possible removal of import duties to make the triangle into a common area in order to facilitate a freer flow of people, goods, and services—decisions which are out of the jurisdiction of the state government (*Straits Times* 1991e).

3.3 Competition between Johor and Singapore

Industrial restructuring, drawn out over the three-country region, can also make competitors as well as partners out of neighbours. In this sense, Singapore's policy of encouraging firms to relocate operations in neighbouring countries is a double-edged sword. The industrial boom in Johor in the latter part of the 1980s is already choking the traditional supply of workers to Singapore (*Business Times* 1989d). The Johor state government has expressed the possibility of stemming the outflow of labour into Singapore if such flows continue at the expense of the state's industrial growth (*Business Times* 1991c). Johor has also recently turned down Singapore's request for more water, arguing that the state's own water consumption over the next thirty years is expected to double because of rapid industrialisation and the accompanying increase in population (*Straits Times* 1991c).

An increase in competition over the next ten years is also likely between the Pasir Gudang seaport in Johor and Singapore port. Malaysia presently uses Singapore for up to 30% of its exports. Every year, over 100 000 TEUs (twenty-foot equivalent units—the average size of a cargo container) of cargo from the Malaysian hinterland flow through Singapore, resulting in a business loss of over M$40 million to the Johor Port Authority by way of port and handling charges (*Business Times* 1990b; *Straits Times* 1991i). However, with the strong Singapore dollar, lower labour costs in Johor (operating cost is estimated to be 70% of that in Singapore), and with the expansion and upgrading of port facilities at Pasir Gudang, more of the Malaysian shipping traffic is likely to go through Johor instead of Singapore.

4 Discussion

Industrialisation and industrial restructuring have led to greater state intervention in order to perpetuate conditions necessary for the valorisation of capital. The Asian NIEs, and Singapore in particular, have radically changed from being low-cost production sites to being areas for the production of higher-value-added goods and services. For Singapore, such a transformation was necessary and has been rapid given the land and labour constraints of a small city-state. However, to continue in this new niche requires greater state intervention in order to maximise labour and land resources. In the case of labour, the state has to attract new sources as well as to conserve existing supplies from lower productive uses.

Such a scenario involves moving to a regional production platform, and, in section 3 of the paper, I outlined the history of the growth-triangle policy as well as analysing the initial areas for cooperation and the potentials for conflict. These potential conflicts arise from the fact that the growth-triangle arrangement is a new geopolitical-economic arrangement that is being created over an existing one. Thus, whether or not this new arrangement can be successfully grafted onto the existing situation depends not only on the three immediate regions of Singapore, Riau, and Johor but also on the impacts that Riau and Johor have in their respective home countries. Economically, the threat of uneven development in relation to the neighbouring regions, and, politically, the fear that these regions will be increasingly drawn into Singapore's orbit, will take some time to erase. As discussed earlier, both threats seem more serious to Johor than to Riau. It is also important to note that the geopolitical-economic situation will change if, on the basis of present and future successes, the triangle of cooperation enlarges to incorporate other regions (in Indonesia and/or Malaysia) and sovereign states (for example, Thailand).

So far, discussion has been focused on efforts to create a broader regional environment conducive for capital. Whether or not these efforts at providing the groundwork prove successful will depend in turn on the location tendencies of industrial enterprises (Perry, 1991). Indeed, one of the major contributions of the new-international-division-of-labour literature has been to detail the far-flung networks of production organised along industrial branches (for example, Dicken, 1986; Henderson, 1989). Will the growth-triangle region matter in the organisation of these networks? Perry (1991) points out that such regional efforts are more likely to protect Singapore's low-value activities rather than the production networks spun by MNCs, because, presumably, the low-value operations are those likely to be handled by local companies wanting to keep operations close to home, whereas MNC networks are less likely to be bounded geographically. The point made by Perry is an important one, because it raises the question of whether MNC investment, local investments, and state investments differ in terms of commitment to particular locations.

Of the three types of capital, MNC capital, organised on a global scale, obviously has most flexibility, whereas state capital, with a political agenda appended onto the logic of the capital, is most committed to the development of particular locations. Local investments, as Perry (1991) suggests, will benefit from the extended region available as a result of regional cooperative efforts, and Singaporean investments in Johor provide evidence of such movements (see table 7). Even if lower value activities benefit, there is no reason why multinational production networks will not respond positively to growth-triangle developments. A region with three different factor endowments and with a convenient movement of people, goods, and services may create trifurcations necessary for developing the Southeast Asian sector of such global networks. But whether MNC capital will stay and allocate production processes within the triangle will depend not only on the attractiveness of the triangle but also on the relative attractiveness of similarly endowed locations in other parts of the world.

However, the different factor endowments of the three partners are not static, and in this sense the common desire for higher-value-added activities will push the partners to become competitors (as in the port-development example discussed earlier), thus tending to erode the differences these countries have in the longer run. Seen in this more dynamic context, the growth-triangle arrangement will buy Singapore some time to sort out its domestic problems with labour and reorganise its priorities for land use before neighbouring regions catch up.

The more specific question concerns whether or not the process of decentralisation created by the triangle will allow Singapore to keep its share of desired activities. Does the regional division of labour created by the growth triangle allow Singapore to move into higher-value-added services *and* manufacturing? There is evidence to support the service–manufacturing division of labour, with Singapore being the regional headquarters of Southeast Asian operations (for example, Ho, 1991). This division is likely to assume more prominence as growth-triangle developments progress (with the completion of more industrial parks, infrastructure, etc), as Singapore, as a city-state with a mature transportation, communication, economic, and social infrastructure, will continue to attract a disproportionate share of services coming into the region.

Whether manufacturing activities in Singapore will be upgraded as a result of the growth triangle is more uncertain. Perry (1991, page 145) suggests, on the basis of the Hong Kong–Guangdong experience, that the decentralisation will lead to an increase in services, but "*moving out* may have mitigated the urgency to *move up* through technological innovation". Such an assumption depends on the type of industrial activity and the strength of state support. In the electronics industry, for example, Singapore, unlike Hong Kong, has moved up the technology ladder with strong state support at a time when both faced the first round of increasing wage costs and labour scarcity (Todd, 1990; UNCTC, 1986). In the process, a network of local material and components suppliers and engineering support has been built up (UNCTC, 1987). Such networks or technical infrastructures are likely at present to keep multinational producers from straying too far from the island.

State support for manufacturing activities will continue to be strong. The deputy prime minister has recently stated the government position on manufacturing as a central element of the Singapore economy for the jobs it creates and its multiplier effects. In line with this position, the government is active to foster the growth of high-value-added industries within the manufacturing sector by ensuring the availability of skilled labour, crucial support services, management and automation equipment, and software (*Straits Times* 1992a). Thus, the established technical base of local suppliers, together with sustained strong state support, may continue to produce the upgrading necessary as well as continue to attract MNC with high-value-added manufacturing activities.

Singapore is presently at a crossroads after years of development; it has acquired the financial and institutional ability to sustain the conditions for accumulation, but, at the same time, is faced with severe labour and land constraints. At this juncture, Singapore, as an NIE operating in a turbulent world with many manufacturing location options, can survive successfully only if state intervention structures local and regional conditions to fit the requirements of international capital.

References

BERI, 1985, "FORCE: Forecast of country environments for strategic planning: Singapore", Business Environment Risk Information Ltd, Suite 10, 1355 Redondo Avenue, Long Beach, CA 90804

Bryant R C, 1989, "The evolution of Singapore as a financial centre", in *Management of Success: The Moulding of Modern Singapore* Eds K S Sandhu, P Wheatley (Institute of Southeast Asian Studies, Singapore)

Business Times 1989a, "Seven workers lost for every eight hired, MNCs report high turnover of staff", 25 January, page 1

1989b, "Inter-industry poaching by some; labour strapped employers lower recruitment criteria", 16 April, page 1

1989c, "Less good but still good, says employers", 23 May, page 12

1989d, "Firms hiring part-time workers to secure more foreign staff", 26 June, page 1

1990a, "Billion S steel mill in the Johor pipeline", 4 March, page 14

1990b, "Johor wants second link with S'pore delayed", 2 May, page 1

1990c, "Muhyiddin floats three growth triangle ideas", 18 November, page 1

1990d, "US concerns triple employment of foreign workers", 11 December, page 2

1991a, "Johor sees surge in S'pore investments", 21 March, page 8

1991b, "DBS Land joins Pernas to built JB township", 7 May, page 3

1991c, "Johor to stem labour flow to Singapore if economy is hurt", 12 June, page 1

1991d, "Stepping stones to a steady water supply", 30 June, page 2

Cohen R B, 1981, "The new international division of labour, multinational corporations and urban hierarchy", in *Urbanization and Urban Planning in Capitalist Society* Eds M Dear, A J Scott (Methuen, New York) pp 287-315

Deyo F (Ed.), 1987 *The Political Economy of the New Asian Industrialism* (Cornell University Press, Ithaca, New York)

Dicken P, 1986 *Global Shift: Industrial Change in a Turbulent World* (Harper and Row, London)

DoS, 1980 *Yearbook of Statistics 1980* Department of Statistics, 8 Shenton Way, #10-01 Treasury Building, Singapore 0106

DoS, 1980-88 *Report on the Census of Industrial Production* Department of Singapore, 8 Shenton Way, #10-01 Treasury Building, Singapore 0106

DoS, 1981 *Census of Population 1980. Singapore Release No.4: Economic Characteristics*, Department of Statistics, 8 Shenton Way, #10-01 Treasury Building, Singapore 0106

DoS, 1983 *Economic and Social Statistics of Singapore 1960-1982*, Department of Statistics, 8 Shenton Way, #10-01 Treasury Building, Singapore 0106

DoS, 1990 *Yearbook of Statistics 1990* Department of Statistics, 8 Shenton Way, #10-01 Treasury Building, Singapore 0106

EDB, various years, "Survey of business expectations of industrial establishments", Economic Development Board, 250 North Bridge Road #24-00, Raffles City Tower, Singapore 0617

Frobel F, Heinrichs J, Kreye O, 1980 *The New International Division of Labour* (Cambridge University Press, Cambridge)

Henderson J, 1986, "The new international division of labour and urban development in the contemporary world system", in *Urbanisation in the DevelopingWorld* Ed. D Drakakis Smith (Croom Helm, Andover, Hants) pp 63-81

Henderson J, 1989 *The Globalisation of High Technology Production* (Routledge, Chapman and Hall, Andover, Hants)

Heyzer N, 1982, "From rural subsistence to an industrial peripheral workforce: an examination of female Malaysian migrants and capital accumulation in Singapore", in *Women and Development: The Sexual Division of Labour in Rural Societies* Ed. L Beneria (Praeger, New York) pp 179-202

Ho K C, 1991, "Studying the city in the new international division of labour", WP-107, Department of Sociology, National University of Singapore, Singapore

Lee H L, 1988, "Managing the Singapore economy", closing address of Global Strategies: The Singapore Partnership, 24-26 October; copy available from Economic Development Board, 250 North Bridge Road #24-00, Raffles City Tower, Singapore 0617

Lee S Y, 1984, "Some aspects of foreign exchange management in Singapore" *Asia Pacific Journal of Management* 1 207-217

Lee S Y, 1986 *The Monetary and Banking Development of Singapore and Malaysia* 2nd edition (Singapore University Press, Singapore)

Lee T Y, 1992, "Regional economic zones in the Asia Pacific: a conceptual overview", unpublished manuscript, available from Institute of Policy Studies, Hon Sui Sen Memorial Library Building, National University of Singapore, Singapore

Kamil Y, Pangestu M, Fredericks C, 1991, "A Malaysian perspective", in *Growth Triangle: The Johor-Singapore-Riau Experience* Ed. T Y Lee (Institute of Southeast Asian Studies, Singapore)

Mirza H, 1986 *Multinationals and the Growth of the Singapore Economy* (Croom Helm, Andover, Hants)

MoL, 1989, "Report on the labour force survey of Singapore 1989", Ministry of Labour, 18 Havelock Road, #06-02, Singapore 0105

MOTI, 1986a, "The Singapore economy: new directions", Economic Committee, Ministry of Trade and Industry, 8 Shenton Way, Treasury Building, Singapore 0106

MOTI, 1986b, "Economic survey of Singapore 1986", Ministry of Trade and Industry, 8 Shenton Way, Treasury Building, Singapore 0106

MOTI, 1991a, "The strategic economic plan: towards a developed nation", Economic Planning Committee, 8 Shenton Way, Treasury Building, Singapore 0106

MOTI, 1991b, "Economic survey of Singapore 1991", Ministry of Trade and Industry, 8 Shenton Way, Treasury Building, Singapore 0106

Natarajan S, Tan J M, 1992 *The Impact of MNC Investments in Malaysia, Singapore and Thailand* (Institute of Southeast Asian Studies, Singapore)

Newsweek 1992, "China's Gold Coast: is Guangdong province too rich for Beijing?", 17 February, pp 8-12

Pang E F, Lim L, 1982, "Foreign labour and economic development in Singapore" *International Migration Review* **16** 548-576

Pangestu M, 1991, "An Indonesian perspective", in *Growth Triangle: The Johor-Singapore-Riau Experience* Ed. T Y Lee (Institute of Southeast Asian Studies, Singapore) pp 75-115

Park S O, 1991, "Government management of industrial change in the Republic of Korea", in *The State and the Spatial Management of Industrial Change* Eds D C Rich, G J R Linge (Routledge, Chapman and Hall, Andover, Hants) pp 74-87

Perry M, 1991, "The Singapore growth triangle: state, capital and labour at a new frontier in the world economy" *Singapore Journal of Tropical Geography* **12** 138-151

Rich D C, Linge G J R, 1991, "The state and industrial change", in *The State and the Spatial Management of Industrial Change* Eds D C Rich, G J R Linge (Routledge, Chapman and Hall, Andover, Hants) pp 1-21

Rodan G, 1989 *The Political Economy of Singapore's Industrialization: National State and International Capital* (Macmillan, London)

SIAA, 1988, "National automation survey 1987/1988", Singapore Industrial Automation Association, 2 Bukit Merah Central #14-03, NPB Building, Singapore 0913

SIAA, 1990, "National automation survey 1989/1990", Singapore Industrial Automation Association, 2 Bukit Merah Central #14-03, NPB Building, Singapore 0913

Straits Times 1987, "Republic among major users of industrial robots", 1 October, page 18

1988a, "Electrical and electronics industry bears brunt of sharp increases in costs of recruitment", 26 January, page 1

1988b, "Lower CPF rates for those above 55", 1 May, page 1

1988c, "Govt gives automation $60 million boost: incentives to help firm reduce dependence on foreign labour", 23 November, page 1

1989a, "Boon or bane?", 6 February, page 10

1989b, "Over-reliance on foreign workers is bad in the long run", 6 February, page 10

1989c, "Lifeline scheme can take in 100,000 in 8 years", 11 July, page 3

1989d, "Part-timers can play a greater role in Singapore", 6 September, page 25

1989e, "Retirement age law may come in two years", 7 September, page 1

1990a, "Johor leads in industrial land sales", 1 February, page 31

1990b, "S'pore-Johor relations have never been better: Muhyiddin", 15 July, page 1

1990c, "Singapore to help Johor develop high-tech park", 17 September, page 16

1990d, "How Johor can be a better neighbour to Singapore", 22 September, page 28

1990e, "Johor proposal to get S'pore's help in high-tech park hailed", 17 October, page 3

Straits Times
1991a, "Johor wants to continue control of its profitable water supply", 14 March, page 15
1991b, "Johor water department to become independent body: MB", 19 March, page 17
1991c, "Singapore requests for more water: 'Johor cannot help'", 1 April, page 16
1991d, "Paper questions KL's plan to run water resources", 28 May, page 14
1991e, "Johor, Riau and Singapore should harmonise customs rates: Dr Habibie", 31 May, page 48
1991f, "Johor wants KL-S'pore joint panel on labour", 28 June, page 22
1991g, "S'pore to help Johor set up manpower training facilities", 24 July, page 3
1991h, "Taiwan's parliament approves S5b steel mill project in Johor", 25 October, page 43
1991i, "KL using S'pore for up to 30% of container exports", 18 November, page 16
1991j, "Federal, Johor teams to meet on growth triangle", 21 December, page 22
1992a, "Manufacturing will remain mainstay of economy: B G Lee", 6 May, page 1
1992b, "Weak S$ will undermine competitiveness: Dr Hu", 8 July, page

Todd D, 1990 *The World Electronics Industry* (Routledge, Chapman and Hall, Andover, Hants)

Turnbull C M, 1977 *A History of Singapore 1819-1975* (Oxford University Press, London)

UNCTC, 1986 *Transnational Corporations in International Semiconductor Industry* United Nations Centre on Transnational Corporations (United Nations, New York)

UNCTC, 1987 *Transnational Corporations and the Electronics Industries of ASEAN Economies* United Nations Centre on Transnational Corporations (United Nations, New York)

Wade R, 1990, "Industrial policy in East Asia: does it lead or follow the market?", in *Manufacturing Miracles* Eds G Gereffi, D L Wyman (Princeton University Press, Princeton, NJ) pp 231-266

18

LIPPINCOTT LIBRARY/
DOCUMENT DELIVERY

WHARTON REPROGRAPHICS

Competitiveness: A Dangerous Obsession

Paul Krugman

THE HYPOTHESIS IS WRONG

IN JUNE 1993, Jacques Delors made a special presentation to the leaders of the nations of the European Community, meeting in Copenhagen, on the growing problem of European unemployment. Economists who study the European situation were curious to see what Delors, president of the EC Commission, would say. Most of them share more or less the same diagnosis of the European problem: the taxes and regulations imposed by Europe's elaborate welfare states have made employers reluctant to create new jobs, while the relatively generous level of unemployment benefits has made workers unwilling to accept the kinds of low-wage jobs that help keep unemployment comparatively low in the United States. The monetary difficulties associated with preserving the European Monetary System in the face of the costs of German reunification have reinforced this structural problem.

It is a persuasive diagnosis, but a politically explosive one, and everyone wanted to see how Delors would handle it. Would he dare tell European leaders that their efforts to pursue economic justice have produced unemployment as an unintended by-product? Would he admit that the EMS could be sustained only at the cost of a recession and face the implications of that admission for European monetary union?

PAUL KRUGMAN is Professor of Economics at the Massachusetts Institute of Technology. His most recent book is *Peddling Prosperity: Economic Sense and Nonsense in the Age of Diminished Expectations* (W. W. Norton).

Reproduced with permission of copyright owner. Further reproduction prohibited.

Guess what? Delors didn't confront the problems of either the welfare state or the EMS. He explained that the root cause of European unemployment was a lack of competitiveness with the United States and Japan and that the solution was a program of investment in infrastructure and high technology.

It was a disappointing evasion, but not a surprising one. After all, the rhetoric of competitiveness—the view that, in the words of President Clinton, each nation is "like a big corporation competing in the global marketplace"—has become pervasive among opinion leaders throughout the world. People who believe themselves to be sophisticated about the subject take it for granted that the economic problem facing any modern nation is essentially one of competing on world markets—that the United States and Japan are competitors in the same sense that Coca-Cola competes with Pepsi—and are unaware that anyone might seriously question that proposition. Every few months a new best-seller warns the American public of the dire consequences of losing the "race" for the 21st century.[1] A whole industry of councils on competitiveness, "geo-economists" and managed trade theorists has sprung up in Washington. Many of these people, having diagnosed America's economic problems in much the same terms as Delors did Europe's, are now in the highest reaches of the Clinton administration formulating economic and trade policy for the United States. So Delors was using a language that was not only convenient but comfortable for him and a wide audience on both sides of the Atlantic.

---

[1] See, for just a few examples, Laura D'Andrea Tyson, *Who's Bashing Whom: Trade Conflict in High-Technology Industries*, Washington: Institute for International Economics, 1992; Lester C. Thurow, *Head to Head: The Coming Economic Battle among Japan, Europe, and America*, New York: Morrow, 1992; Ira C. Magaziner and Robert B. Reich, *Minding America's Business: The Decline and Rise of the American Economy*, New York: Vintage Books, 1983; Ira C. Magaziner and Mark Patinkin, *The Silent War: Inside the Global Business Battles Shaping America's Future*, New York: Vintage Books, 1990; Edward N. Luttwak, *The Endangered American Dream: How to Stop the United States from Becoming a Third World Country and How to Win the Geo-economic Struggle for Industrial Supremacy*, New York: Simon and Schuster, 1993; Kevin P. Phillips, *Staying on Top: The Business Case for a National Industrial Strategy*, New York: Random House, 1984; Clyde V. Prestowitz, Jr., *Trading Places: How We Allowed Japan to Take the Lead*, New York: Basic Books, 1988; William S. Dietrich, *In the Shadow of the Rising Sun: The Political Roots of American Economic Decline*, University Park: Pennsylvania State University Press, 1991; Jeffrey E. Garten, *A Cold Peace: America, Japan, Germany, and the Struggle for Supremacy*, New York: Times Books, 1992; and Wayne Sandholtz et al., *The Highest Stakes: The Economic Foundations of the Next Security System*, Berkeley Roundtable on the International Economy (BRIE), Oxford University Press, 1992.

Unfortunately, his diagnosis was deeply misleading as a guide to what ails Europe, and similar diagnoses in the United States are equally misleading. The idea that a country's economic fortunes are largely determined by its success on world markets is a hypothesis, not a necessary truth; and as a practical, empirical matter, that hypothesis is flatly wrong. That is, it is simply not the case that the world's leading nations are to any important degree in economic competition with each other, or that any of their major economic problems can be attributed to failures to compete on world markets. The growing obsession in most advanced nations with international competitiveness should be seen, not as a well-founded concern, but as a view held in the face of overwhelming contrary evidence. And yet it is clearly a view that people very much want to hold—a desire to believe that is reflected in a remarkable tendency of those who preach the doctrine of competitiveness to support their case with careless, flawed arithmetic.

This article makes three points. First, it argues that concerns about competitiveness are, as an empirical matter, almost completely unfounded. Second, it tries to explain why defining the economic problem as one of international competition is nonetheless so attractive to so many people. Finally, it argues that the obsession with competitiveness is not only wrong but dangerous, skewing domestic policies and threatening the international economic system. This last issue is, of course, the most consequential from the standpoint of public policy. Thinking in terms of competitiveness leads, directly and indirectly, to bad economic policies on a wide range of issues, domestic and foreign, whether it be in health care or trade.

MINDLESS COMPETITION

MOST PEOPLE who use the term "competitiveness" do so without a second thought. It seems obvious to them that the analogy between a country and a corporation is reasonable and that to ask whether the United States is competitive in the world market is no

Reproduced with permission of copyright owner. Further reproduction prohibited.

different in principle from asking whether General Motors is competitive in the North American minivan market.

A trade surplus may be a sign of national weakness, a deficit a sign of strength.

In fact, however, trying to define the competitiveness of a nation is much more problematic than defining that of a corporation. The bottom line for a corporation is literally its bottom line: if a corporation cannot afford to pay its workers, suppliers, and bondholders, it will go out of business. So when we say that a corporation is uncompetitive, we mean that its market position is unsustainable—that unless it improves its performance, it will cease to exist. Countries, on the other hand, do not go out of business. They may be happy or unhappy with their economic performance, but they have no well-defined bottom line. As a result, the concept of national competitiveness is elusive.

One might suppose, naively, that the bottom line of a national economy is simply its trade balance, that competitiveness can be measured by the ability of a country to sell more abroad than it buys. But in both theory and practice a trade surplus may be a sign of national weakness, a deficit a sign of strength. For example, Mexico was forced to run huge trade surpluses in the 1980s in order to pay the interest on its foreign debt since international investors refused to lend it any more money; it began to run large trade deficits after 1990 as foreign investors recovered confidence and began to pour in new funds. Would anyone want to describe Mexico as a highly competitive nation during the debt crisis era or describe what has happened since 1990 as a loss in competitiveness?

Most writers who worry about the issue at all have therefore tried to define competitiveness as the combination of favorable trade performance and something else. In particular, the most popular definition of competitiveness nowadays runs along the lines of the one given in Council of Economic Advisors Chairman Laura D'Andrea Tyson's *Who's Bashing Whom?:* competitiveness is "our ability to produce goods and services that meet the test of international competition while our citizens enjoy a standard of living that is both rising and sustainable." This sounds reasonable. If you think about it, however, and test your thoughts against the facts, you will find out that there is much less to this definition than meets the eye.

Consider, for a moment, what the definition would mean for an economy that conducted very little international trade, like the United States in the 1950s. For such an economy, the ability to balance its trade is mostly a matter of getting the exchange rate right. But because trade is such a small factor in the economy, the level of the exchange rate is a minor influence on the standard of living. So in an economy with very little international trade, the growth in living standards—and thus "competitiveness" according to Tyson's definition—would be determined almost entirely by domestic factors, primarily the rate of productivity growth. That's domestic productivity growth, period—not productivity growth relative to other countries. In other words, for an economy with very little international trade, "competitiveness" would turn out to be a funny way of saying "productivity" and would have nothing to do with international competition.

But surely this changes when trade becomes more important, as indeed it has for all major economies? It certainly could change. Suppose that a country finds that although its productivity is steadily rising, it can succeed in exporting only if it repeatedly devalues its currency, selling its exports ever more cheaply on world markets. Then its standard of living, which depends on its purchasing power over imports as well as domestically produced goods, might actually decline. In the jargon of economists, domestic growth might be outweighed by deteriorating terms of trade.[2] So "competitiveness" could

---

[2] An example may be helpful here. Suppose that a country spends 20 percent of its income on imports, and that the prices of its imports are set not in domestic but in foreign currency. Then if the country is forced to devalue its currency—reduce its value in foreign currency—by 10 percent, this will raise the price of 20 percent of the country's spending basket by 10 percent, thus raising the overall price index by 2 percent. Even if domestic *output* has not changed, the country's real *income* will therefore have fallen by 2 percent. If the country must repeatedly devalue in the face of competitive pressure, growth in real income will persistently lag behind growth in real output.

It's important to notice, however, that the size of this lag depends not only on the amount of devaluation but on the share of imports in spending. A 10 percent devaluation of the dollar against the yen does not reduce U.S. real income by 10 percent—in fact, it reduces U.S. real income by only about 0.2 percent because only about 2 percent of U.S. income is spent on goods produced in Japan.

Reproduced with permission of copyright owner. Further reproduction prohibited.

turn out really to be about international competition after all.

There is no reason, however, to leave this as a pure speculation; it can easily be checked against the data. Have deteriorating terms of trade in fact been a major drag on the U.S. standard of living? Or has the rate of growth of U.S. real income continued essentially to equal the rate of domestic productivity growth, even though trade is a larger share of income than it used to be?

To answer this question, one need only look at the national income accounts data the Commerce Department publishes regularly in the *Survey of Current Business.* The standard measure of economic growth in the United States is, of course, real GNP—a measure that divides the value of goods and services produced in the United States by appropriate price indexes to come up with an estimate of real national output. The Commerce Department also, however, publishes something called "command GNP." This is similar to real GNP except that it divides U.S. exports not by the export price index, but by the price index for U.S. imports. That is, exports are valued by what Americans can buy with the money exports bring. Command GNP therefore measures the volume of goods and services the U.S. economy can "command"—the nation's purchasing power—rather than the volume it produces.[3] And as we have just seen, "competitiveness" means something different from "productivity" if and only if purchasing power grows significantly more slowly than output.

Well, here are the numbers. Over the period 1959-73, a period of vigorous growth in U.S. living standards and few concerns about international competition, real GNP per worker-hour grew 1.85 percent annually, while command GNP per hour grew a bit faster, 1.87 percent. From 1973 to 1990, a period of stagnating living standards, command GNP growth per hour slowed to 0.65 percent. Almost all (91 percent) of that slowdown, however, was explained by a decline in domestic productivity growth: real GNP per hour grew only 0.73 percent.

---

[3] In the example in the previous footnote, the devaluation would have no effect on real GNP, but command GNP would have fallen by two percent. The finding that in practice command GNP has grown almost as fast as real GNP therefore amounts to saying that events like the hypothetical case in footnote one are unimportant in practice.

Countries do not compete with each other the way corporations do.

Similar calculations for the European Community and Japan yield similar results. In each case, the growth rate of living standards essentially equals the growth rate of domestic productivity—not productivity relative to competitors, but simply domestic productivity. Even though world trade is larger than ever before, national living standards are overwhelmingly determined by domestic factors rather than by some competition for world markets.

How can this be in our interdependent world? Part of the answer is that the world is not as interdependent as you might think: countries are nothing at all like corporations. Even today, U.S. exports are only 10 percent of the value-added in the economy (which is equal to GNP). That is, the United States is still almost 90 percent an economy that produces goods and services for its own use. By contrast, even the largest corporation sells hardly any of its output to its own workers; the "exports" of General Motors—its sales to people who do not work there—are virtually all of its sales, which are more than 2.5 times the corporation's value-added.

Moreover, countries do not compete with each other the way corporations do. Coke and Pepsi are almost purely rivals: only a negligible fraction of Coca-Cola's sales go to Pepsi workers, only a negligible fraction of the goods Coca-Cola workers buy are Pepsi products. So if Pepsi is successful, it tends to be at Coke's expense. But the major industrial countries, while they sell products that compete with each other, are also each other's main export markets and each other's main suppliers of useful imports. If the European economy does well, it need not be at U.S. expense; indeed, if anything a successful European economy is likely to help the U.S. economy by providing it with larger markets and selling it goods of superior quality at lower prices.

International trade, then, is not a zero-sum game. When productivity rises in Japan, the main result is a rise in Japanese real wages; American or European wages are in principle at least as likely to rise as to fall, and in practice seem to be virtually unaffected.

It would be possible to belabor the point, but the moral is clear:

Reproduced with permission of copyright owner. Further reproduction prohibited.

while competitive problems could arise in principle, as a practical, empirical matter the major nations of the world are not to any significant degree in economic competition with each other. Of course, there is always a rivalry for status and power—countries that grow faster will see their political rank rise. So it is always interesting to *compare* countries. But asserting that Japanese growth diminishes U.S. status is very different from saying that it reduces the U.S. standard of living—and it is the latter that the rhetoric of competitiveness asserts.

One can, of course, take the position that words mean what we want them to mean, that all are free, if they wish, to use the term "competitiveness" as a poetic way of saying productivity, without actually implying that international competition has anything to do with it. But few writers on competitiveness would accept this view. They believe that the facts tell a very different story, that we live, as Lester Thurow put it in his best-selling book, *Head to Head*, in a world of "win-lose" competition between the leading economies. How is this belief possible?

CARELESS ARITHMETIC

ONE OF THE REMARKABLE, startling features of the vast literature on competitiveness is the repeated tendency of highly intelligent authors to engage in what may perhaps most tactfully be described as "careless arithmetic." Assertions are made that sound like quantifiable pronouncements about measurable magnitudes, but the writers do not actually present any data on these magnitudes and thus fail to notice that the actual numbers contradict their assertions. Or data are presented that are supposed to support an assertion, but the writer fails to notice that his own numbers imply that what he is saying cannot be true. Over and over again one finds books and articles on competitiveness that seem to the unwary reader to be full of convincing evidence but that strike anyone familiar with the data as strangely, almost eerily inept in their handling of the numbers. Some examples can best illustrate this point. Here are three cases of careless arithmetic, each of some interest in its own right.

Trade Deficits and the Loss of Good Jobs. In a recent article published in Japan, Lester Thurow explained to his audience the importance of reducing the Japanese trade surplus with the United States. U.S. real wages, he pointed out, had fallen six percent during the Reagan and Bush years, and the reason was that trade deficits in manufactured goods had forced workers out of high-paying manufacturing jobs into much lower-paying service jobs.

This is not an original view; it is very widely held. But Thurow was more concrete than most people, giving actual numbers for the job and wage loss. A million manufacturing jobs have been lost because of the deficit, he asserted, and manufacturing jobs pay 30 percent more than service jobs.

Both numbers are dubious. The million-job number is too high, and the 30 percent wage differential between manufacturing and services is primarily due to a difference in the length of the workweek, not a difference in the hourly wage rate. But let's grant Thurow his numbers. Do they tell the story he suggests?

The key point is that total U.S. employment is well over 100 million workers. Suppose that a million workers were forced from manufacturing into services and as a result lost the 30 percent manufacturing wage premium. Since these workers are less than 1 percent of the U.S. labor force, this would reduce the average U.S. wage rate by less than 1/100 of 30 percent—that is, by less than 0.3 percent.

This is too small to explain the 6 percent real wage decline *by a factor of* 20. Or to look at it another way, the annual wage loss from deficit-induced deindustrialization, which Thurow clearly implies is at the heart of U.S. economic difficulties, is on the basis of his own numbers roughly equal to what the U.S. spends on health care every week.

Something puzzling is going on here. How could someone as intelligent as Thurow, in writing an article that purports to offer hard quantitative evidence of the importance of international competition to the U.S. economy, fail to realize that the evidence he offers clearly shows that the channel of harm that he identifies was *not* the culprit?

High Value-added Sectors. Ira Magaziner and Robert Reich, both now influential figures in the Clinton Administration, first reached a broad audience with their 1982 book, *Minding America's Business.* The book advocated a U.S. industrial policy, and in the introduction the authors offered a seemingly concrete quantitative basis for such a policy: "Our

Reproduced with permission of copyright owner. Further reproduction prohibited.

standard of living can only rise if (i) capital and labor increasingly flow to industries with high value-added per worker and (ii) we maintain a position in those industries that is superior to that of our competitors."

Economists were skeptical of this idea on principle. If targeting the right industries was simply a matter of moving into sectors with high value-added, why weren't private markets already doing the job?[4] But one might dismiss this as simply the usual boundless faith of economists in the market; didn't Magaziner and Reich back their case with a great deal of real-world evidence?

Well, *Minding America's Business* contains a lot of facts. One thing it never does, however, is actually justify the criteria set out in the introduction. The choice of industries to cover clearly implied a belief among the authors that high value-added is more or less synonymous with high technology, but nowhere in the book do any numbers compare actual value-added per worker in different industries.

Value Added Per Worker, 1988
(in thousands of dollars)

CIGARETTES	488
PETROLEUM REFINING	283
AUTOS	99
STEEL	97
AIRCRAFT	68
ELECTRONICS	64
ALL MANUFACTURING	66

Such numbers are not hard to find. Indeed, every public library in America has a copy of the *Statistical Abstract of the United States,* which each year contains a table presenting value-added and employment by industry in U.S. manufacturing. All one needs to do, then, is spend a few minutes in the library with a calculator to come up with a table that ranks U.S. industries by value-added per worker.

The table on this page shows selected entries from pages 740–744 of the 1991 *Statistical Abstract.* It turns out that the U.S. industries with really high value-added per worker are in sectors with very high ratios of capital to labor, like cigarettes and petroleum refining. (This was predictable: because capital-intensive industries must earn a normal return on large investments, they must charge prices that are a larger markup over labor costs than labor-intensive industries, which means that they

---

[4] "Value-added" has a precise, standard meaning in national income accounting: the value added of a firm is the dollar value of its sales, minus the dollar value of the inputs it purchases from other firms, and as such it is easily measured. Some people who use the term, however, may be unaware of this definition and simply use "high value-added" as a synonym for "desirable."

Competitiveness advocates are eerily inept in their handling of the numbers.

have high value-added per worker). Among large industries, value-added per worker tends to be high in traditional heavy manufacturing sectors like steel and autos. High-technology sectors like aerospace and electronics turn out to be only roughly average.

This result does not surprise conventional economists. High value-added per worker occurs in sectors that are highly capital-intensive, that is, sectors in which an additional dollar of capital buys little extra value-added. In other words, there is no free lunch.

But let's leave on one side what the table says about the way the economy works, and simply note the strangeness of the lapse by Magaziner and Reich. Surely they were not calling for an industrial policy that would funnel capital and labor into the steel and auto industries in preference to high-tech. How, then, could they write a whole book dedicated to the proposition that we should target high value-added industries without ever checking to see which industries they meant?

Labor Costs. In his own presentation at the Copenhagen summit, British Prime Minister John Major showed a chart indicating that European unit labor costs have risen more rapidly than those in the United States and Japan. Thus he argued that European workers have been pricing themselves out of world markets.

But a few weeks later Sam Brittan of the *Financial Times* pointed out a strange thing about Major's calculations: the labor costs were not adjusted for exchange rates. In international competition, of course, what matters for a U.S. firm are the costs of its overseas rivals measured in dollars, not marks or yen. So international comparisons of labor costs, like the tables the Bank of England routinely publishes, always convert them into a common currency. The numbers presented by Major, however, did not make this standard adjustment. And it was a good thing for his presentation that they didn't. As Brittan pointed out, European labor costs have not risen in relative terms when the exchange rate adjustment is made.

If anything, this lapse is even odder than those of Thurow or Mag-

Reproduced with permission of copyright owner. Further reproduction prohibited.

aziner and Reich. How could John Major, with the sophisticated statistical resources of the U.K. Treasury behind him, present an analysis that failed to make the most standard of adjustments?

These examples of strangely careless arithmetic, chosen from among dozens of similar cases, by people who surely had both the cleverness and the resources to get it right, cry out for an explanation. The best working hypothesis is that in each case the author or speaker wanted to believe in the competitive hypothesis so much that he felt no urge to question it; if data were used at all, it was only to lend credibility to a predetermined belief, not to test it. But why are people apparently so anxious to define economic problems as issues of international competition?

THE THRILL OF COMPETITION

THE COMPETITIVE metaphor—the image of countries competing with each other in world markets in the same way that corporations do—derives much of its attractiveness from its seeming comprehensibility. Tell a group of businessmen that a country is like a corporation writ large, and you give them the comfort of feeling that they already understand the basics. Try to tell them about economic concepts like comparative advantage, and you are asking them to learn something new. It should not be surprising if many prefer a doctrine that offers the gain of apparent sophistication without the pain of hard thinking. The rhetoric of competitiveness has become so widespread, however, for three deeper reasons.

First, competitive images are exciting, and thrills sell tickets. The subtitle of Lester Thurow's huge best-seller, *Head to Head*, is "The Coming Economic Battle among Japan, Europe, and America"; the jacket proclaims that "the decisive war of the century has begun . . . and America may already have decided to lose." Suppose that the subtitle had described the real situation: "The coming struggle in which each big economy will succeed or fail based on its own efforts, pretty much independently of how well the others do." Would Thurow have sold a tenth as many books?

Second, the idea that U.S. economic difficulties hinge crucially on our failures in international competition somewhat paradoxically makes those difficulties seem easier to solve. The productivity of the average American worker is determined by a complex array of factors, most of them unreachable by any likely government policy. So if you accept the reality that our "competitive" problem is really a domestic productivity problem pure and simple, you are unlikely to be optimistic about any dramatic turnaround. But if you can convince yourself that the problem is really one of failures in international competition—that imports are pushing workers out of high-wage jobs, or subsidized foreign competition is driving the United States out of the high value-added sectors—then the answers to economic malaise may seem to you to involve simple things like subsidizing high technology and being tough on Japan.

Finally, many of the world's leaders have found the competitive metaphor extremely useful as a political device. The rhetoric of competitiveness turns out to provide a good way either to justify hard choices or to avoid them. The example of Delors in Copenhagen shows the usefulness of competitive metaphors as an evasion. Delors had to say something at the EC summit; yet to say anything that addressed the real roots of European unemployment would have involved huge political risks. By turning the discussion to essentially irrelevant but plausible-sounding questions of competitiveness, he bought himself some time to come up with a better answer (which to some extent he provided in December's white paper on the European economy—a paper that still, however, retained "competitiveness" in its title).

By contrast, the well-received presentation of Bill Clinton's initial economic program in February 1993 showed the usefulness of competitive rhetoric as a motivation for tough policies. Clinton proposed a set of painful spending cuts and tax increases to reduce the Federal deficit. Why? The real reasons for cutting the deficit are disappointingly undramatic: the deficit siphons off funds that might otherwise have been productively invested, and thereby exerts a steady if small drag on U.S. economic growth. But Clinton was able instead to offer a stirring patriotic appeal, calling on the nation to act now in order to make the economy competitive in the global market—with the implication that dire economic consequences would follow if the United States does not.

Reproduced with permission of copyright owner. Further reproduction prohibited.

Many people who know that "competitiveness" is a largely meaningless concept have been willing to indulge competitive rhetoric precisely because they believe they can harness it in the service of good policies. An overblown fear of the Soviet Union was used in the 1950s to justify the building of the interstate highway system and the expansion of math and science education. Cannot the unjustified fears about foreign competition similarly be turned to good, used to justify serious efforts to reduce the budget deficit, rebuild infrastructure, and so on?

A few years ago this was a reasonable hope. At this point, however, the obsession with competitiveness has reached the point where it has already begun dangerously to distort economic policies.

THE DANGERS OF OBSESSION

THINKING AND SPEAKING in terms of competitiveness poses three real dangers. First, it could result in the wasteful spending of government money supposedly to enhance U.S. competitiveness. Second, it could lead to protectionism and trade wars. Finally, and most important, it could result in bad public policy on a spectrum of important issues.

During the 1950s, fear of the Soviet Union induced the U.S. goverment to spend money on useful things like highways and science education. It also, however, led to considerable spending on more doubtful items like bomb shelters. The most obvious if least worrisome danger of the growing obsession with competitiveness is that it might lead to a similar misallocation of resources. To take an example, recent guidelines for government research funding have stressed the importance of supporting research that can improve U.S. international competitiveness. This exerts at least some bias toward inventions that can help manufacturing firms, which generally compete on international markets, rather than service producers, which generally do not. Yet most of our employment and value-added is now in services, and lagging productivity in services rather than manufactures has been the single most important factor in the stagnation of U.S. living standards.

A much more serious risk is that the obsession with competitiveness will lead to trade conflict, perhaps even to a world trade war.

Most of those who have preached the doctrine of competitiveness have not been old-fashioned protectionists. They want their countries to win the global trade game, not drop out. But what if, despite its best efforts, a country does not seem to be winning, or lacks confidence that it can? Then the competitive diagnosis inevitably suggests that to close the borders is better than to risk having foreigners take away high-wage jobs and high-value sectors. At the very least, the focus on the supposedly competitive nature of international economic relations greases the rails for those who want confrontational if not frankly protectionist policies.

We can already see this process at work, in both the United States and Europe. In the United States, it was remarkable how quickly the sophisticated interventionist arguments advanced by Laura Tyson in her published work gave way to the simple-minded claim by U.S. Trade Representative Mickey Kantor that Japan's bilateral trade surplus was costing the United States millions of jobs. And the trade rhetoric of President Clinton, who stresses the supposed creation of high-wage jobs rather than the gains from specialization, left his administration in a weak position when it tried to argue with the claims of NAFTA foes that competition from cheap Mexican labor will destroy the U.S. manufacturing base.

Perhaps the most serious risk from the obsession with competitiveness, however, is its subtle indirect effect on the quality of economic discussion and policymaking. If top government officials are strongly committed to a particular economic doctrine, their commitment inevitably sets the tone for policy-making on all issues, even those which may seem to have nothing to do with that doctrine. And if an economic doctrine is flatly, completely and demonstrably wrong, the insistence that discussion adhere to that doctrine inevitably blurs the focus and diminishes the quality of policy discussion across a broad range of issues, including some that are very far from trade policy per se.

Consider, for example, the issue of health care reform, undoubtedly the most important economic initiative of the Clinton administration, almost surely an order of magnitude more important to U.S. living standards than anything that might be done about trade

Reproduced with permission of copyright owner. Further reproduction prohibited.

Competitiveness risks distorting the quality of domestic economic policy.

policy (unless the United States provokes a full-blown trade war). Since health care is an issue with few direct international linkages, one might have expected it to be largely insulated from any distortions of policy resulting from misguided concerns about competitiveness.

But the administration placed the development of the health care plan in the hands of Ira Magaziner, the same Magaziner who so conspicuously failed to do his homework in arguing for government promotion of high value-added industries. Magaziner's prior writings and consulting on economic policy focused almost entirely on the issue of international competition, his views on which may be summarized by the title of his 1990 book, *The Silent War*. His appointment reflected many factors, of course, not least his long personal friendship with the first couple. Still, it was not irrelevant that in an administration committed to the ideology of competitiveness Magaziner, who has consistently recommended that national industrial policies be based on the corporate strategy concepts he learned during his years at the Boston Consulting Group, was regarded as an economic policy expert.

We might also note the unusual process by which the health care reform was developed. In spite of the huge size of the task force, recognized experts in the health care field were almost completely absent, notably though not exclusively economists specializing in health care, including economists with impeccable liberal credentials like Henry Aaron of the Brookings Institution. Again, this may have reflected a number of factors, but it is probably not irrelevant that anyone who, like Magaziner, is strongly committed to the ideology of competitiveness is bound to have found professional economists notably unsympathetic in the past—and to be unwilling to deal with them on any other issue.

To make a harsh but not entirely unjustified analogy, a government wedded to the ideology of competitiveness is as unlikely to make good economic policy as a government committed to creationism is to make good science policy, even in areas that have no direct relationship to the theory of evolution.

ADVISERS WITH NO CLOTHES

IF THE OBSESSION with competitiveness is as misguided and damaging as this article claims, why aren't more voices saying so? The answer is, a mixture of hope and fear.

On the side of hope, many sensible people have imagined that they can appropriate the rhetoric of competitiveness on behalf of desirable economic policies. Suppose that you believe that the United States needs to raise its savings rate and improve its educational system in order to raise its productivity. Even if you know that the benefits of higher productivity have nothing to do with international competition, why not describe this as a policy to enhance competitiveness if you think that it can widen your audience? It's tempting to pander to popular prejudices on behalf of a good cause, and I have myself succumbed to that temptation.

As for fear, it takes either a very courageous or very reckless economist to say publicly that a doctrine that many, perhaps most, of the world's opinion leaders have embraced is flatly wrong. The insult is all the greater when many of those men and women think that by using the rhetoric of competitiveness they are demonstrating their sophistication about economics. This article may influence people, but it will not make many friends.

Unfortunately, those economists who have hoped to appropriate the rhetoric of competitiveness for good economic policies have instead had their own credibility appropriated on behalf of bad ideas. And somebody has to point out when the emperor's intellectual wardrobe isn't all he thinks it is.

So let's start telling the truth: competitiveness is a meaningless word when applied to national economies. And the obsession with competitiveness is both wrong and dangerous.

Reproduced with permission of copyright owner. Further reproduction prohibited.

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

19

LIPPINCOTT LIBRARY
DOCUMENT DELIVERY

WHARTON REPROGRAPHICS

Response

Proving My Point

Paul Krugman

My article in the March/April issue of *Foreign Affairs* has obviously upset many people. Some of my critics claim that I misrepresented their position, that despite their insistent use of the word "competitiveness" they have never believed that the major industrial nations are engaged in a competitive economic struggle. Others claim that I have gotten the economics wrong: that countries *are* engaged in a competitive struggle. Indeed, some of them make both claims in the same response.

MOVING TARGET

Lester C. Thurow vigorously denies ever asserting that international competition is a central issue for the U.S. economy. In particular, he cites page counts from his 1985 book, *The Zero-Sum Solution*, to demonstrate that domestic factors are his principal concern. But Thurow's most recent book is *Head to Head*, which follows its provocative title with the subtitle, *The Coming Battle Among Japan, Europe and America*. The book jacket asserts that the "decisive war of the century has begun." The text asserts over and over that the major economic powers are now engaged in "win-lose" competition for world markets, a competition that has taken the place of the military competition between East and West. Thurow now says that international strategic competition is no more than seven percent of the problem; did the typical reader of *Head to Head* get this message?

Similarly, Stephen S. Cohen denies that he, or indeed anyone else with whom I should "deign to take difference," has ever said the things I claim competitiveness advocates believe. But in 1987 Cohen, together with John Zysman, published *Manufacturing Matters*, a book that seemed to say two (untrue) things: the long-term downward trend in the share of manufacturing in U.S. employment is largely due to foreign competition, and this declining share is a major economic problem.

After their initial denial, both Cohen and Thurow proceed to argue that international competition is of crucial importance after all. In this they are joined by

---

PAUL KRUGMAN is Professor of Economics at the Massachusetts Institute of Technology.

Reproduced with permission of copyright owner. Further reproduction prohibited.

Clyde V. Prestowitz, Jr., who at least makes no bones about believing that trade and trade policy are the central issue for the U.S. economy. Does Cohen believe that Prestowitz—or James Fallows, who expressed similar views in his new book, *Looking at the Sun*—is one of those people with whom I should not deign to argue?

SLOPPY MATH: PART II

Of all the elements in my article, the section on careless arithmetic—the strange pattern of errors in reporting or using data in articles and books on competitiveness—has enraged the most people. Both Thurow and Prestowitz have taken care to fill their responses with a blizzard of numbers and calculations. However, some of the numbers are puzzling.

For example, Thurow says that imports are 14 percent of U.S. GDP, while exports are only 10 percent, and that reducing imports to equal exports would add $250 billion to the sales of U.S. manufacturers. But according to *Economic Indicators*, the monthly statistical publication of the Joint Economic Committee, U.S. imports in 1993 were only 11.4 percent of GDP, while exports were 10.4 percent. Even the current account deficit, a broader measure that includes some additional debit items, was only $109 billion. If the United States were to cut imports by $250 billion, far from merely balancing its trade as Thurow asserts, the United States would run a current account *surplus* of $140 billion—that is, more than the 2 percent maximum of GDP U.S. negotiators have demanded Japan set as a target!

Or consider Prestowitz, who derides my claim that high-technology industries, commonly described as "high value" sectors, actually have much lower value added per worker than traditional "high volume," heavy industrial sectors. I have aggregated too much by looking at broad sectors like electronics, he says; I should look at the highest-tech lines of business, like semiconductors, where value added per worker is $234,000. Prestowitz should report the results of his research to the Department of Commerce, whose staff has obviously incorrectly calculated (in the *Annual Survey of Manufactures*) that in 1989 value added per worker in Standard Industrial Classification 3674 (semiconductors and related devices) was $96,487—closer to the $76,709 per worker in SIC 2096 (potato chips and related snacks) than to the $187,569 in SIC 3711 (motor vehicles and car bodies).[1]

Everyone makes mistakes, although it is surprising when men who are supposed to be experts on international competition do not have even a rough idea of the size of the U.S. trade deficit or know how to look up a standard industrial statistic. The interesting point, however, is that the mistakes made by Thurow, Prestowitz and other competitiveness advocates are not random errors; they are always biased in the same direction. That is, the advocates always err in a direction that makes international competition seem more important than it really is.

---

[1] I don't know why Thurow thinks the U.S. trade deficit is four times as big as it actually is. I have, however, tracked down Prestowitz's number. It is not value added per employee; it is shipments (which are always larger than value added) divided by the number of production workers (who are only a fraction of total employment, especially in high-technology industries).

Beyond these petty, if revealing, errors of fact are a series of conceptual misunderstandings. For example, Prestowitz argues that productivity in sectors that compete on world markets is much more important than productivity in non-traded service sectors because the former determine wage rates throughout the economy. For example, because U.S. manufacturing workers are much more productive than their Third World counterparts, U.S. barbers, who do not have a comparable productivity advantage, also get high wages. But Prestowitz fails to notice that the converse is also true: service productivity affects the real wages of manufacturing workers. Because the high relative productivity of U.S. manufacturing is not matched in the haircut sector, haircuts by those well-paid barbers are much more expensive than haircuts in the Third World; as a result real wages of U.S. manufacturing workers (that is, wages in terms of what they can buy, including haircuts) are not as high as they would be if U.S. barbers were more productive. With careful thought, one realizes that real wages depend on the overall productivity of the economy, with no special presumption that productivity in manufacturing—or in internationally traded sectors in general—deserves any more attention or active promotion than productivity elsewhere.

THE BETTMANN ARCHIVE

Friedrich List, the new cult figure

Cohen makes essentially the same mistake when he complains that I underestimated the effects of competitive pressure because I focused only on import and export prices and did not consider the further impacts of that pressure on profits and wages. He somehow fails to realize that a change in wages or profits that is not reflected in import or export prices cannot change overall U.S. real income—it can only redistribute profits to one group within the United States at the expense of another. That is why the effect of international price competition on U.S. real income can be measured by the change in the ratio of export to import prices—full stop. And the effects of changes in this ratio on the U.S. economy have, as I showed in my article, been small.

Or consider Thurow's analysis of the benefits that would accrue to the United States if it could roll back imports (leaving aside the inaccuracy of his numbers). He asserts that the United States could create five million new jobs in import-competing sectors, and he assumes that all five million jobs represent a net addi-

Reproduced with permission of copyright owner. Further reproduction prohibited.

tion to employment. But this assumption is unrealistic. As this reply was being written, the Federal Reserve was raising interest rates in an effort to rein in a recovery that it feared would proceed too far, that is, lead to excessive employment, producing a renewed surge in inflation. Some people think that the Fed is tightening too soon, but the essential point is that the growth of employment is not determined by the ability of the United States to sell goods on world markets or to compete with imports, but by the Fed's judgement of what will not set off inflation. So suppose that the United States were to impose import quotas, adding millions of jobs in import-competing sectors. The Fed would respond by raising interest rates to prevent an overheated economy, and most if not all of the job gains would be matched by job losses elsewhere.

THINGS ADD UP

In each of these cases, my critics seem to have forgotten the most basic principle of economics: things add up. Higher employment in import-competing industries must come either through a reduction in unemployment, in which case one must ask whether the implied unemployment rate (about three percent in Thurow's example) is feasible, or at the expense of jobs elsewhere in the economy, in which case no overall job gain takes place. If higher manufacturing wages lead to a higher wage rate for barbers without higher tonsorial productivity, the gain must come at someone else's expense. Since it is hard to see how foreigners pay for more expensive American haircuts, that wage gain can only redistribute the benefits of manufacturing productivity from one set of American workers to another, not increase the total gains. In their haste to assign great importance to international competition, my critics, like the inventors of perpetual motion machines, have failed to realize that there are conservation principles that any story about the economy must honor.

But perhaps Cohen, Thurow and Prestowitz stumble on economic basics because they are so eager to get to their main point, which is that advanced economic theory, and in particular the theory of strategic trade policy, supports their obsession with competitiveness.

Prestowitz's central assertion is that the theory of strategic trade policy, which he for some reason thinks I invented in a paper about aircraft competition (the actual inventors were James Brander and Barbara Spencer, who never mentioned aircraft), justifies aggressively interventionist trade policies. He further asserts that economists in general, and I in particular, have run away from that implication for ideological reasons.

Well, that's not quite the real story. It is true that in the early 1980s professional economists became aware that one of the implications of new theories of international trade was a possible role for strategic policies to promote exports in certain industries. Confronted with a new idea that was exciting, potentially important but untested, these economists began a sustained process of research, probing the weak points, confronting the new idea with the data. After all, lots of things could be true in principle. For example in certain theoretical situations a tax cut could definitely stimulate the economy so much that government revenues would actually rise, and it would be very nice if that were the actual situation; but unfortunately it isn't. Similarly, it is definitely possible to imagine a situation in which, because of all of the market imperfections Thurow dwells on, a clever strategic trade policy would sharply raise U.S. real income. And it would be very nice if the United States could devise such a policy. But is that possibility really there? To answer that question requires looking hard at the facts.

And so over the course of the last ten years a massive international research program has explored the prospects for strategic trade policy.[2] Two broad conclusions emerge. First, to identify which industries should receive strategic promotion or the appropriate form and level of promotion is very difficult. Second, the payoffs of even a successful strategic trade policy are likely to be very modest—certainly far less even than Thurow's "seven percent solution," which is closer to the entire share of international trade in the U.S. economy.

Research results are always open to challenge, especially in an inexact field like economics. If Prestowitz wants to point out specific failings in the dozens of painstaking empirical studies of strategic trade that have been carried out over the past decade, by all means let him do so. His remarks about the subject, however, strongly suggest that while he is happy to mention strategic trade theory in support of his policy writing, Prestowitz has not read any of the economic literature.

I do, however, agree with Prestowitz on one point. More people should read the works of Friedrich List. If they do, they may wonder why this turgid, confused writer—whose theory led him to predict that Holland and Denmark would be condemned to permanent economic backwardness unless they sought political union with Germany—has suddenly become a favorite of Fallows, Prestowitz and others. The new cult of List bears an uncanny resemblance to the right-wing supply-siders' canonization of the classical French economist Jean-Baptiste Say, who claimed that the economy as a whole could never suffer from the falls in aggregate demand that produce recessions.[3] The motive of the supply-siders was, of course, to cover simplistic ideas with a veneer of faux scholarship.

In contrast to Prestowitz and Thurow, who offer coherent if flawed reasons to worry about international competition,

---

[2] The original paper on strategic trade policy was James Brander and Barbara Spencer, "Export Subsidies and International Market Share Rivalry," *Journal of International Economics*, February 1985, pp. 83-100. See also Paul Krugman, ed., *Strategic Trade Policy and the New International Economics*, Cambridge: MIT Press, 1986; Robert Feenstra, ed., *Empirical Methods for International Trade*, Chicago: University of Chicago Press, 1988; Robert Baldwin, ed., *Trade Policy Issues and Empirical Analysis*, Chicago: University of Chicago Press, 1988; and Paul Krugman and Alasdair Smith, eds., *Empirical Studies of Strategic Trade Policy*, Chicago: University of Chicago Press, 1994.

[3] Fallows officially elevated List to guru status in his article "How the World Works," *The Atlantic Monthly*, December 1993, pp. 60-87. Readers may wish to compare the elevation of Say by Jude Wanniski in his influential supply-side tract, *The Way the World Works*, New York: Basic Books, 1978.

Reproduced with permission of copyright owner. Further reproduction prohibited.

Cohen offers a more difficult target. Basically, he asks us to accept "competitiveness" as a kind of ineffable essence that cannot be either defined or measured. Data that seem to suggest the importance of this essence are cited as "indicators," whatever that means, while those that do not are dismissed as unreliable. Both in his article and other writings he has persistently used a rhetoric that seems to portray international trade as a game with winners and losers, but when challenged on any particular point he denies having said it. I guess I don't understand how a concept so elusive can be a useful guide to policy.

My original article in *Foreign Affairs* argued that a doctrine that views world trade as a competitive struggle has become widely accepted, that this view is wrong but that there is nonetheless an intense desire to believe in that doctrine. The article enraged many, especially when it asserted that the desire to believe in competitive struggle repeatedly leads highly intelligent authors into surprising lapses in their handling of concepts and data. I could not, however, have asked for a better demonstration of my point than the responses published in this issue. ⊕

Call for any book

Immediate shipment worldwide

Amex, MasterCard, Visa or check

All books in print available

Ask about our overnight gift delivery nationwide

Gift wrapping available

Mail orders welcome

Open 24 hours every day

1-800-255-2665

In CT or Worldwide
(203) 966-5470
FAX 1-203-966-4329

BOOK CALL
59 Elm Street
New Canaan, CT 06840

Foreign Affairs (ISSN 0015-7120), July/August 1994, Volume 73, Number 4. Published six times annually (January, March, May, July, September, November) at 58 East 68th Street, New York, NY 10021. Subscriptions: U.S., $38.00; Canada, $47.00; other countries via air, $68.00 per year. Second class postage paid in New York, NY, and at additional mailing offices. Postmaster: Send address changes to *Foreign Affairs*, P.O. Box 420235, Palm Coast, FL 32142-0235.

Reproduced with permission of copyright owner. Further reproduction prohibited.

Review of Economics and Statistics
Volume 79, Issue 3, 1997

20

LIPPINCOTT LIBRAF
DOCUMENT DELIVEF

WHARTON REPROGRAPHICS

TRADE BARRIERS AND TRADE FLOWS ACROSS COUNTRIES AND INDUSTRIES

Jong-Wha Lee and Phillip Swagel*

Abstract—We use disaggregated data on trade flows, production, and trade barriers for 41 countries in 1988 to examine the political and economic determinants of nontariff barriers, as well as the impact of protection (both tariff and nontariff) on trade flows. We use an econometric framework that allows for the simultaneous determination of trade barriers and trade flows. Our results are consistent with political-economy theories of the determinants of protection: even after accounting for industry- and country-specific factors, nations tend to protect industries that are weak, in decline, politically important, or threatened by import competition, but provide less protection to industries in which exports are important.

I. Introduction

THEORETICAL interest has recently focused on the determinants of nations' trade barriers. Underlying these theories is the implicit belief that there are common economic and political factors that can explain the structure of protection across countries and industries.

In this paper we use disaggregated cross-country, cross-industry data for production, trade barriers, and trade flows of manufactured goods in 1988 to examine the political and economic determinants of global nontariff barriers (NTBs). As tariff levels have fallen and remained bound by General Agreement on Tariffs and Trade (GATT) strictures, NTBs have increasingly become the instrument of choice for protection. The calls for protection from import-competing industries indicate that the pattern of trade is likely to have an effect on the structure of protection. Since protection (both tariff and nontariff) clearly affects trade flows, we use an econometric framework that allows for the simultaneous determination of trade barriers and trade flows. Unlike previous studies, our sample includes both developed countries with low barriers and developing countries with substantial protection across all manufacturing industries. A novelty of this paper is combining disaggregated data on production, trade flows, and trade barriers for a broad range of countries.

Our results are consistent with political-economy theories of the determination of trade protection. We find that NTBs are determined by more than just industry- or country-specific factors. Nations tend to protect weak industries, as well as industries in decline, but provide less protection to export industries, perhaps out of fear of foreign retaliation. Large industries, which we think of as being politically important, also receive protection in the form of NTBs. We find that tariffs are used in conjunction with NTBs rather than as a substitute means of protection, and that NTB protection is endogenous in the sense that these barriers are in part determined by industry conditions such as import penetration. Finally, we find that trade barriers reduce imports, though our results are somewhat mixed as to the relative impact of tariffs versus NTBs.

The paper proceeds as follows. Section II discusses some previous work on the political economy of trade protection, as well as empirical studies of the effects of protection on trade flows. In section III we specify a model of trade barrier determination, along with a model of trade flows based on the monopolistic competition model of trade. Section IV describes the data, after which we present our empirical results in section V. Section VI concludes.

II. The Political Economy of Trade Protection

Given the economic consensus regarding the efficiency of free trade, models of trade barrier determination typically turn to political-economy explanations.[1] The key insight is that the structure of the political system can be important in the determination of trade protection when a distinction exists between consumers and producers, or between different groups of consumers or producers. The benefits of trade protection typically accrue to a narrow group of stakeholders in the protected industry, whereas the costs are spread over a much larger number of consumers, each of whom loses only a small amount. This asymmetry means that protection may be politically efficient even if it is inefficient in an economic sense.

Baldwin (1982) discusses these political motivations for trade protection, whereas Magee et al. (1989) and Rodrik (1994) survey a variety of models in which political factors influence trade policy. Bhagwati (1982) also contains several interesting models of the determinants of protection. Grossman and Helpman (1994) provide a more rigorous theoretical foundation for this literature through an explicit model of the process by which different interest groups bid for protection. With perfect competition in the product market, their model predicts that the structure of protection depends on two factors: the elasticity of import demand, which indicates the degree to which trade barriers distort welfare, and the ratio of imports to domestic output, which reflects the political importance of the domestic industry. The surveys mentioned above discuss numerous other contributions to this literature.

There is by now a substantial body of empirical work on the political economy of trade protection. The papers by

Received for publication January 20, 1995. Revision accepted for publication November 3, 1995.

* Korea University and NBER, and International Monetary Fund, respectively.

We thank Judith Dean, Jim Harrigan, Judy Hellerstein, Mike Knetter, Kala Krishna, Jim Rauch, Joe Stone, Jerry Thursby, and Andrew Warner for helpful discussions; seminar participants at the 1994 Econometric Society Winter Meetings, the Federal Reserve Board, the New York Fed, the Fletcher School of Law and Diplomacy, and Purdue University for comments; and Refik Erzan and Will Martin for assistance in obtaining data. We have benefited from suggestions made by two referees of this *Review*.

---

[1] There is of course a large literature on strategic trade policy, but this typically examines trade barriers in a particular industry, rather than the determination of the entire structure of trade protection.

College and the Massachusetts Institute of Technology

Review of Economics and Statistics
Volume 79, Issue 3, 1997

LIPPINCOTT LIBRARY
DOCUMENT DELIVERY
20
WHARTON REPROGRAPHICS

TRADE BARRIERS AND TRADE FLOWS ACROSS COUNTRIES AND INDUSTRIES

Jong-Wha Lee and Phillip Swagel*

Abstract—We use disaggregated data on trade flows, production, and trade barriers for 41 countries in 1988 to examine the political and economic determinants of nontariff barriers, as well as the impact of protection (both tariff and nontariff) on trade flows. We use an econometric framework that allows for the simultaneous determination of trade barriers and trade flows. Our results are consistent with political-economy theories of the determinants of protection: even after accounting for industry- and country-specific factors, nations tend to protect industries that are weak, in decline, politically important, or threatened by import competition, but provide less protection to industries in which exports are important.

I. Introduction

THEORETICAL interest has recently focused on the determinants of nations' trade barriers. Underlying these theories is the implicit belief that there are common economic and political factors that can explain the structure of protection across countries and industries.

In this paper we use disaggregated cross-country, cross-industry data for production, trade barriers, and trade flows of manufactured goods in 1988 to examine the political and economic determinants of global nontariff barriers (NTBs). As tariff levels have fallen and remained bound by General Agreement on Tariffs and Trade (GATT) strictures, NTBs have increasingly become the instrument of choice for protection. The calls for protection from import-competing industries indicate that the pattern of trade is likely to have an effect on the structure of protection. Since protection (both tariff and nontariff) clearly affects trade flows, we use an econometric framework that allows for the simultaneous determination of trade barriers and trade flows. Unlike previous studies, our sample includes both developed countries with low barriers and developing countries with substantial protection across all manufacturing industries. A novelty of this paper is combining disaggregated data on production, trade flows, and trade barriers for a broad range of countries.

Our results are consistent with political-economy theories of the determination of trade protection. We find that NTBs are determined by more than just industry- or country-specific factors. Nations tend to protect weak industries, as well as industries in decline, but provide less protection to export industries, perhaps out of fear of foreign retaliation. Large industries, which we think of as being politically important, also receive protection in the form of NTBs. We find that tariffs are used in conjunction with NTBs rather than as a substitute means of protection, and that NTB protection is endogenous in the sense that these barriers are in part determined by industry conditions such as import penetration. Finally, we find that trade barriers reduce imports, though our results are somewhat mixed as to the relative impact of tariffs versus NTBs.

The paper proceeds as follows. Section II discusses some previous work on the political economy of trade protection, as well as empirical studies of the effects of protection on trade flows. In section III we specify a model of trade barrier determination, along with a model of trade flows based on the monopolistic competition model of trade. Section IV describes the data, after which we present our empirical results in section V. Section VI concludes.

II. The Political Economy of Trade Protection

Given the economic consensus regarding the efficiency of free trade, models of trade barrier determination typically turn to political-economy explanations.[1] The key insight is that the structure of the political system can be important in the determination of trade protection when a distinction exists between consumers and producers, or between different groups of consumers or producers. The benefits of trade protection typically accrue to a narrow group of stakeholders in the protected industry, whereas the costs are spread over a much larger number of consumers, each of whom loses only a small amount. This asymmetry means that protection may be politically efficient even if it is inefficient in an economic sense.

Baldwin (1982) discusses these political motivations for trade protection, whereas Magee et al. (1989) and Rodrik (1994) survey a variety of models in which political factors influence trade policy. Bhagwati (1982) also contains several interesting models of the determinants of protection. Grossman and Helpman (1994) provide a more rigorous theoretical foundation for this literature through an explicit model of the process by which different interest groups bid for protection. With perfect competition in the product market, their model predicts that the structure of protection depends on two factors: the elasticity of import demand, which indicates the degree to which trade barriers distort welfare, and the ratio of imports to domestic output, which reflects the political importance of the domestic industry. The surveys mentioned above discuss numerous other contributions to this literature.

There is by now a substantial body of empirical work on the political economy of trade protection. The papers by

---

Received for publication January 20, 1995. Revision accepted for publication November 3, 1995.

* Korea University and NBER, and International Monetary Fund, respectively.

We thank Judith Dean, Jim Harrigan, Judy Hellerstein, Mike Knetter, Kala Krishna, Jim Rauch, Joe Stone, Jerry Thursby, and Andrew Warner for helpful discussions; seminar participants at the 1994 Econometric Society Winter Meetings, the Federal Reserve Board, the New York Fed, the Fletcher School of Law and Diplomacy, and Purdue University for comments; and Refik Erzan and Will Martin for assistance in obtaining data. We have benefited from suggestions made by two referees of this *Review*.

[1] There is of course a large literature on strategic trade policy, but this typically examines trade barriers in a particular industry, rather than the determination of the entire structure of trade protection.

© 1997 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology

Marvel and Ray (1983, 1987, 1981a,b, 1985) examine various aspects of the implication of the theoretical literature that the structure of protection across industries depends on the particular political and economic characteristics of each industry. Ray (1981b) estimates equations for the simultaneous determination of imports and trade barriers (both tariff and nontariff) in the United States, and finds that NTBs in the United States fell mainly on capital-intensive, low-skill industries. Ray (1981a) estimates trade and protection equations for both the United States and an aggregate of foreign countries. While he finds that tariffs and NTBs were used as complements, he finds surprisingly little effect of trade protection on U.S. imports. More recently, Trefler (1993) estimates trade and NTB equations for the United States and shows that taking into account the simultaneous determination of imports and trade protection results in a substantially larger estimate of the effects of protection on imports. Trefler also finds that political factors and proxies for industrial structure, such as measures of union density and industry concentration, have the expected positive impact on the level of protection in the United States.

Marvel and Ray (1983) estimate equations for the determination of U.S. tariffs and NTBs alone. They find that protection was given to politically important industries and industries under threat from imports, whereas healthy industries received less protection. They ascribe this to the Peltzman–Becker theory of regulation, which suggests that policy makers will seek to share an industry's good fortune with weaker sectors. Ray and Marvel (1985) estimate tariff and NTB equations alone for the United States, Canada, Japan, and the European Community (EC) as a whole. They find broad similarities in the structure of protection in these countries. Although tariff rates were generally low, they find that NTBs were used to undercut this apparent liberality, particularly in the EC. Dick (1994) also finds that NTBs were used to compensate industries affected by reduced tariffs, whereas Gawande (1995) finds that NTBs in the United States are largely used to retaliate against foreign protection. Caves (1976) and Helleiner (1977) provide econometric analysis of the determinants of Canadian tariffs, whereas Dean et al. (1994) provide a case-study analysis of recent trade liberalizations in a number of developing countries.

Leamer (1990) provides evidence on the extent and effect of tariff and NTBs in Organization for Economic Cooperation and Development (OECD) countries in 1983, focusing particularly on the effect of these barriers on Latin American countries. Harrigan (1993) examines the effect of protection on trade flows between OECD countries.

Like Ray (1981a,b) and Trefler (1993), we combine the literatures on the determination of trade barriers and trade flows, and attempt to control for their simultaneous determination. One advantage of our analysis is that our sample includes both developed and developing countries, and thus encompasses substantially more variation in the structure of trade flows and trade barriers.

III. Models of Trade Barriers and Trade Flows

A. *Determinants of Trade Barriers*

As discussed in section II, the structure of production is probably best thought of as determined by both economic and political factors. To take into account the notion that the political power of the industry is likely to be important, we use two proxies for a sector's political influence: the size of the industry as measured by its share of value added within a country, as well as the industry's share of labor. Of course, these are likely to be imperfect indicators of political importance, since small industries may be seen as crucial to national security, or in many countries, might be directly or indirectly owned by policy makers.

There may also be political pressure to protect "weak" industries, such as those with low wages or productivity. To examine this, we include wage per worker and value added per worker in our model of trade barrier determination. Including both variables allows us to examine the extent to which protection is given to industries in which wages are high conditional on industry productivity—that is, do nations provide more protection to industries with labor rents? Of course, it may be impossible to make inferences about causality here, since protection could lead to a "lazy" industry with low productivity or high wages rather than a weak industry receiving protection.

To examine the tendency of declining industries to receive protection, we include the five-year change in the log wage per worker (from 1982 to 1987) as an explanatory variable. If there is profit sharing in an industry, declining wages would indicate declining rents, and thus shifting comparative advantage. We would also expect political pressure for protection to be stronger in industries "threatened" by import competition. Counterbalancing this, however, might be a desire to avoid retaliation by other nations aimed at domestic export industries. To examine these possibilities, we include as explanatory variables both a measure of import penetration (the share of imports in domestic use) as well as a measure of the importance of industry exports (the share of exports in the output of each industry).

Not surprisingly, there are other determinants of trade protection for which we could not obtain data. For example, Grossman and Helpman (1994) show that trade barriers are more likely to exist the lower the own price elasticity of demand for an industry's product, since this entails a smaller deadweight loss to consumers. Similarly, the higher the foreign price elasticity of supply, the more effective will be a given trade barrier in changing the pattern of trade. Since we could not obtain cross-country, cross-industry data on elasticities and industry characteristics such as concentration ratios, we must instead rely on the inclusion of industry fixed effects to account for any omitted industry-specific effects. This will work to the extent that these omitted factors are constant across countries. We also present specifications that include country fixed effects. These will account for differences in the overall trade regime, which are specific to particular nations.

We take tariff rates as exogenous determinants of NTBs. While not strictly correct, this is probably not too bad an assumption relative to NTBs, since tariff rates in many countries are under GATT strictures. For the United States, Ray (1981b) finds no feedback from NTBs to tariffs.

In our basic specification, then, NTBs in industry i of country j are expected to respond to sectoral imports and other economic and political factors,

$$NTB_{ij} = f(IM_{ij}/DU_{ij}, \Delta(W_{ij}/L_{ij}), X_{ij}/Q_{ij}, W_{ij}/L_{ij}, VA_{ij}/L_{ij}, VA_{ij}/VA_j, tariff_{ij}, BMP_j) \tag{1}$$

where

IM_{ij}/DU_{ij} = import penetration (imports as share of domestic use)
$\Delta(W_{ij}/L_{ij})$ = five-year change of real wage per worker
X_{ij}/Q_{ij} = exports per domestic production
W_{ij}/L_{ij} = wage per worker
VA_{ij}/L_{ij} = labor productivity (value added per worker)
VA_{ij}/VA_j = sectoral share of value added
$tariff_{ij}$ = tariff rate
BMP_j = black-market premium.

As discussed above, wages and labor productivity proxy for each sector's competitive position, while the sectoral share of value added (or, alternatively, the sectoral share of labor L_{ij}/L_j) is meant as a proxy of political power. The change of real wages captures the evolution of each industry. Since declining industries typically call for protection, we would expect this to be associated negatively with NTBs. Tariff rates and the black-market premium are included in the regression to examine whether different varieties of trade restrictions tend to be used as substitutes or in tandem.

We specify equation (1) using a logarithmic functional form. As detailed in the next section, our measure of NTBs is a coverage ratio which is bounded from below at zero. To take this censoring into account, we specify the NTB equation as a tobit,

$$\begin{aligned}\log(1 + NTB^*_{ij}) = c_0 &+ c_1 \log(IM_{ij}/DU_{ij}) \\ &+ c_2 \Delta \log(W_{ij}/L_{ij}) \\ &+ c_3 \log(X_{ij}/Q_{ij}) + c_4 \log(W_{ij}/L_{ij}) \\ &+ c_5 \log(VA_{ij}/L_{ij}) \\ &+ c_6 \log(VA_{ij}/VA_j) \\ &+ c_7 \log(1 + tariff_{ij}) \\ &+ c_8 \log(1 + BMP_j) + u_{Nij}\end{aligned} \tag{2}$$

$$NTB_{ij} = \begin{cases} NTB^*_{ij}, & \text{if } NTB^*_{ij} > 0 \\ 0, & \text{otherwise.}\end{cases}$$

B. Monopolistic Competition Model of Trade

For the determination of trade flows, we use a simple monopolistic competition model of trade based on Helpman and Krugman (1985). With identical homothetic preferences for consumers, each country consumes identical proportions of each product. Goods are assumed to be imperfectly substitutable and differentiated by country of origin. Since the production of each variety of a product occurs in only one country, the model gives a prediction of the volume of trade as follows:

$$IM_{ij} = s_j(Q_i - Q_{ij}) \tag{3}$$

where

IM_{ij} = import of good i by country j
Q_{ij} = production of good i in country j
Q_i = total world production of good i
s_j = share of country j in world income.

Equation (3) states that country j's import of good i is proportional to the amount of good i produced outside country j. One prediction from equation (3) is that the import share of a good in a particular country is inversely related to that country's share of production of world output of that good,

$$IM_{ij}/DU_{ij} = G(Q_{ij}/Q_i), \qquad G' < 0. \tag{4}$$

Domestic use DU_{ij} equals production plus imports minus exports. Equation (4) provides a basic framework with which to estimate the volume and composition of trade.[2]

The monopolistic competition model gives a prediction of the volume of trade in the absence of trade barriers. When the Helpman–Krugman model is extended to include trade policies, as in Flam and Helpman (1987), it gives ambiguous predictions about the effects of protection on welfare, production, and trade flows. In the simple framework above, however, the model unambiguously predicts that the presence of trade barriers, such as tariffs, NTBs, and exchange controls, will diminish the volume of trade.

We augment equation (4) to include measures of trade barriers and transport costs, and we adopt a logarithmic functional form,

$$\begin{aligned}\log(IM_{ij}/DU_{ij}) = a_0 &+ a_1 \log(Q_{ij}/Q_i) \\ &+ a_2 \log(distance_j) \\ &+ a_3 \log(1 + tariff_{ij}) \\ &+ a_4 \log(1 + NTB_{ij}) \\ &+ a_5 \log(1 + BMP_j) + u_{Iij}.\end{aligned} \tag{5}$$

---

[2] Some examples include Helpman (1984), Lawrence (1987), Bergstrand (1989), and Harrigan (1993).

The variables *tariff* and *NTB* measure the intensity of trade barriers on good i in country j, while the black-market premium *BMP* is meant to capture the distortionary effects of exchange controls that might hinder imports. These are described in section IV. A trade-weighted measure of distance between the capital of each country and the capitals of its trading partners is used to proxy for transportation costs, as in Bergstrand (1985, 1989). Our specification for trade flows is similar to that of Harrigan (1993), who estimates the effects of tariff and NTBs (which are taken as exogenous) on bilateral trade flows in manufactures in OECD countries. As with the NTB equation (2), we estimate specifications with industry fixed effects, as well as with both industry and country fixed effects.

As pointed out by Leamer and Bowen (1981) and Leamer (1988), the response of imports to tariffs and NTBs is likely to vary substantially across industries, since this depends on elasticities of supply and demand, which might differ widely across industries. In contrast, there is likely to be less variation in these elasticities within a given industry across countries, so that estimating cross-country regressions separately for each industry is preferable to pooling across industries. Unfortunately, as we discuss in the next section, we have data on trade barriers for only one year, and our data are further limited to each nation's total trade flows in each industry, so that we must pool across both countries and industries to obtain sufficient degrees of freedom for estimation. The results for coefficients a_3 and a_4 in equation (5) should thus be interpreted as average values of the effects of tariffs and NTBs on imports.

Since the production of each good is determined simultaneously with trade flows, we follow Harrigan (1995) and use factor endowments to instrument for sectoral production. We use the economywide factor endowments of each nation's capital stock, labor force, human capital, and land area. These are also described in section IV. Unfortunately data on sector-specific inputs are not available for our wide range of countries. An immediate implication of this is that we cannot compare our results with those from a Heckscher–Ohlin model like the one Harkness (1978) estimates for the United States, in which factor endowments determine the pattern of trade.

The production share and the distance measure are used as instruments for imports and thus do not appear in the NTB equation, whereas the "political-economy" variables instrument for the level of protection, and do not enter into the import volume equation. This latter identifying restriction is clearly more troubling, since trade flows might have a direct effect on productivity, the industry share of value added (or workers), and the evolution of wages.

We assume that the error terms u_I and u_N are distributed with a bivariate normal, and estimate equations (2) and (5) jointly using the simultaneous-equations tobit methodology of Nelson and Olson (1978). We also estimate the equations using the corresponding single-equation methodologies (tobit and ordinary least squares), and calculate a Hausman test of the null hypothesis of no simultaneity bias.

IV. Data

A. *Trade and Protection Data*

To measure the degree of trade barriers across industry and country, we use the data set of trade control measures (TCMs) compiled by UNCTAD (1991). This provides information on both tariffs and additional charges on imports in 1988, as well as information on the coverage of nontariff measures (NTMs) at the most detailed level of the Customs Cooperation Council Nomenclature (CCCN)—four digits plus up to two alphabetic codes.

The tariff provided by UNCTAD is the ad valorem nominal rate for total import charges; this includes all duties and customs fees collected at national borders. The measure of NTBs reports the coverage ratio for "core" NTMs; this includes essentially all nontariff restrictions applied at the border, including quantitative restrictions (QRs), voluntary export restraints (VERs), and advance payment requirements. Note, however, that the measures of both tariffs and NTBs do not include restrictions which apply inside national borders, such as consumption taxes in countries with no domestic production. See UNCTAD (1987, 1991) and Laird and Yeats (1990) for details.

The coverage ratio indicates the extent to which the tariff lines within a CCCN category are affected by core NTMs. For instance, the index equals zero for a particular four- or five-digit CCCN category if no NTMs apply to any of the products that make up that category. The CCCN category for autos might include tariff lines for both small and medium-sized products. If a country has an NTM on small but not medium-sized cars, then the coverage ratio for that CCCN category would equal 0.5, regardless of the composition of auto imports. The NTM coverage ratio thus captures only the frequency of the nontariff restrictions, but provides no information on the severity of the distortions or the distribution of the resulting quota rents.

The data set also provides the value of trade flows taken from the United Nations COMTRADE database. The trade flow data are at the level of disaggregation needed to match the tariff and NTM data (either four or five digits of SITC revision 2), so that the UNCTAD database facilitates combining the trade barrier data with import values.

Unfortunately, reliable cross-country, cross-industry production data are available only at the three-digit level of the ISIC classification, so we must aggregate up the data on trade barriers and trade flows by weighting them by the country's import value. Weighting by the own import values has the well-known problem that a high level of protection typically results in a low level of imports, and thus a low weight. As a check on this bias, table 1 reports tariffs and NTBs aggregated from the three-digit level to a single value for each country (what might be thought of as the "zero-

TABLE 1.—MEASURES OF PROTECTION BY COUNTRY

		Tariff Rate			Nontariff Barriers		
Country	Sectors	Import Weighted	Production Weighted	Standard Deviation	Import Weighted	Production Weighted	Standard Deviation
United States	27	4.9	4.5	4.2	37.5	26.6	29.9
Canada	27	7.4	7.5	5.1	4.1	6.0	12.7
Sweden	27	4.3	3.6	3.5	7.7	8.4	21.0
Hong Kong	25	0.0	0.0	0.0	0.0	0.0	0.1
Germany, West	27	7.4	7.1	6.0	22.3	15.1	27.4
Japan	26	4.9	3.5	7.7	11.3	6.9	11.0
Denmark	27	7.1	7.2	4.1	18.2	21.0	27.0
France	27	7.4	7.1	14.3	18.4	18.7	26.1
United Kingdom	27	7.1	7.1	6.8	17.9	18.3	27.8
Belgium	27	7.0	7.0	7.6	19.6	29.5	28.2
Italy	22	7.6	7.6	14.4	20.9	23.7	27.1
Netherlands	27	7.1	7.1	7.6	20.6	21.4	28.1
Singapore	20	0.9	0.6	1.8	1.2	1.1	1.7
Spain	27	6.8	6.9	4.8	13.9	17.7	22.1
Ireland	25	7.5	7.5	8.4	20.8	29.0	27.0
Barbados	18	20.7	17.7	13.7	33.9	38.7	32.1
Cyprus	23	19.0	26.5	18.7	49.0	53.3	36.2
Venezuela	27	24.9	33.3	24.4	0.5	3.2	23.8
Greece	27	7.0	8.5	8.6	25.5	34.3	25.8
Portugal	24	7.1	7.1	3.4	19.1	22.7	20.2
Korea	26	16.8	17.7	13.1	1.8	3.3	14.9
Mauritius	22	65.2	88.5	91.5	37.5	43.3	15.2
Malaysia	26	11.7	16.1	12.5	6.8	7.7	8.6
Syria	25	22.0	26.6	27.7	61.0	61.7	25.0
Chile	27	26.7	28.8	14.9	8.8	15.3	23.5
Turkey	27	27.9	26.0	28.2	90.4	84.8	29.4
Colombia	27	45.6	50.4	22.5	64.1	74.8	29.7
Thailand	27	35.2	46.4	21.5	5.5	13.7	20.0
Ecuador	27	38.4	50.3	27.0	51.5	72.3	28.9
Jordan	25	30.9	26.1	18.5	25.6	22.5	31.1
Tunisia	15	27.7	34.0	37.4	63.7	66.5	31.2
Sri Lanka	24	39.5	58.6	35.0	14.7	39.5	32.3
Guatemala	27	14.8	26.5	26.7	100.0	98.8	9.6
Egypt	26	14.7	16.9	20.0	38.5	46.8	32.1
Indonesia	23	17.7	21.7	16.7	10.7	6.6	11.4
Philippines	26	27.2	26.8	13.4	61.1	53.2	32.8
Pakistan	26	41.4	44.0	34.0	9.7	19.3	34.1
Zimbabwe	26	26.3	26.7	11.8	96.0	96.1	20.5
Bangladesh	25	61.9	76.5	50.4	57.5	65.9	30.0
India	26	152.5	153.4	45.2	94.5	93.4	11.6
Kenya	20	30.7	33.9	21.6	23.9	31.1	35.2

Note: Sorted by 1988 real GDP per capita from Summers and Heston (1991), adjusted for the terms of trade.

digit" level), using both import and production weights. Just as import weights will understate trade barriers, production weights will overstate them, since a high trade barrier will result in larger domestic production than would occur in the absence of all barriers. At this "zero-digit" level, the simple correlation between the two weighting schemes is 0.982 for tariffs and 0.971 for NTBs, whereas the rank correlations are 0.966 for tariffs and 0.945 for NTBs. These high values provide some hope that using import weights to go from four or five digits to three digits will not introduce too much bias into the measures of trade protection.

B. *Production, Labor, and Wage Data*

Data on gross output and value added as well as industry wages and employment at the three-digit level of the ISIC classification system are from the United Nations Industrial Statistics Yearbook, volume 1, as found in the BESD database of the World Bank. The data on wages include wages, salaries, and supplements. These wage data, along with the data for gross output and value added are in home-country currency. To match the trade flow data, which are reported in U.S. dollars, we convert the currencies using the exchange rate series from the Summers and Heston (1991) data set. Implicit gross domestic product (GDP) deflators from the Summers and Heston data set are used to deflate lagged wages in order to obtain the change in wage per worker in 1988 dollars.

The data for the black-market premium, distance, and the factor endowments of land, area, human capital, and labor force are from Barro and Lee (1993). The measure of distance is the import-weighted distance between a nation's capital and the capitals of its trading partners. The black-market premium on foreign exchange is measured as an average of the period from 1980 to 1984.

C. *Features of Trade Barriers*

Table 1 gives a summary of protection by country, whereas table 2 summarizes protection across industries. Our sample is limited by the availability of both the trade data and the production data. For example, the UNCTAD trade barrier database does not include non-EU developed countries such as Austria, Switzerland, Australia, and New Zealand. Similarly, reliable production data are not available for many countries, particularly developing countries.

Construction of the import- and production-weighted tariffs and NTBs are discussed above; the standard deviations shown in tables 1 and 2 are of the unweighted tariff rates and NTM coverage ratios. As expected, import-weighting typically results in smaller measures of protection than production weighting although, again, the two are highly correlated.

Even a brief glance at table 1 reveals the dramatically higher levels of protection in developing countries than in developed. (The table is sorted by 1988 per-capita real GDP from the Summers and Heston data set.) Indeed, numbering the countries with 1 being the richest (United States) and 41 the poorest (Kenya), we find rank correlation coefficients of 0.464 between the income ordering (1–41) and import-weighted NTBs, and 0.793 between the income rank and import-weighted tariffs. On the other hand, while tariff rates are quite low in most developed countries, these nations employ a notably higher level of NTB protection (relative to their tariff levels)—there is a rank correlation of −0.289 between the income rank and the ratio of NTBs to tariffs. This is consistent with the findings of Marvel and Ray (1983) and Dick (1994) that NTBs were used to offset the diminished tariffs negotiated in the various GATT rounds. The higher tariff levels in developing countries probably also reflect the greater importance of these relatively easily collected revenues in government finance.

Table 2 shows that protection, particularly NTBs, is concentrated in certain industries, notably food, clothing, steel, and transport equipment. NTB protection similarly appears to be given to large industries—the rank correlation is 0.45 between import-weighted NTBs and the industry share of world output. In contrast, the analogous rank correlation for tariffs is only 0.07.

Of course, the makeup of the above list of industries covered by NTBs is for the most part explained by industry-specific managed trade arrangements, such as the Multi-Fibre Arrangement for textiles and clothing, and the web of bilateral quantitative restraints that govern trade in steel and automobiles. This suggests that the pattern of protection Hufbauer and Rosen (1986) ascribe to the United States might apply to the rest of the world as well—that protection is "special" in that it is industry specific. If this were the case, then including industry fixed effects in the basic regression specification of equation (5) would eliminate the statistical significance of political-economic determinants, such as the political importance or competitive position of an industry in each particular country.

TABLE 2.—MEASURES OF PROTECTION BY INDUSTRY

				Tariff Rates			Nontariff Barriers		
ISIC Code	Industry	Share of Total Output	Observations	Import Weighted	Production Weighted	Standard Deviation	Import Weighted	Production Weighted	Standard Deviation
311	Food products	0.120	40	9.8	11.6	33.6	45.9	37.2	30.0
313	Beverages	0.019	39	5.7	10.4	84.6	43.1	36.8	37.8
314	Tobacco	0.008	38	39.7	25.5	68.2	2.3	5.5	40.5
321	Textiles	0.039	39	11.7	15.9	24.5	69.8	63.4	38.2
322	Apparel, except footwear	0.019	35	12.3	17.2	30.4	71.7	57.6	35.7
323	Leather products	0.004	38	6.1	12.9	31.3	1.4	7.7	40.0
324	Footwear, except rubber/plastic	0.003	36	13.3	20.3	44.5	33.8	23.2	41.9
331	Wood products, except furniture	0.019	39	5.0	5.5	30.3	0.5	1.0	42.1
332	Furniture, except metal	0.012	40	6.4	7.7	46.2	0.9	10.1	46.5
341	Paper and products	0.036	41	6.5	4.5	22.0	1.7	1.8	34.3
342	Printing and publishing	0.042	40	6.5	4.9	22.7	1.2	0.7	29.8
351	Industrial chemicals	0.057	41	10.2	9.9	23.5	9.1	10.1	30.1
352	Other chemicals	0.043	40	8.0	9.9	24.2	7.4	11.7	31.3
353	Petroleum refineries	0.039	37	7.9	11.1	26.9	2.3	5.8	40.0
354	Misc. petroleum and coal	0.004	28	1.0	7.5	19.8	1.6	7.6	33.1
355	Rubber products	0.012	39	9.3	14.5	32.2	4.9	8.6	35.5
361	Pottery, china, earthenware	0.002	37	9.4	12.4	41.5	0.6	7.4	43.6
362	Glass and products	0.007	39	10.1	9.6	29.0	6.0	8.8	34.5
369	Other nonmetal mineral products	0.023	40	6.9	8.2	26.3	10.2	13.1	34.8
371	Iron and steel	0.046	36	9.8	12.2	38.0	47.7	41.1	34.6
372	Nonferrous metals	0.022	34	6.4	6.7	24.0	1.4	2.9	35.1
381	Fabricated metal products	0.056	41	6.3	5.3	27.7	3.9	4.4	32.7
382	Nonelectrical machinery	0.102	39	7.0	4.7	23.8	10.5	12.2	32.0
383	Machinery electric	0.102	41	8.6	5.8	28.3	14.3	10.1	33.3
384	Transport equipment	0.132	40	7.9	6.1	30.7	25.5	27.0	38.8
385	Professional and scientific equipment	0.021	35	6.5	4.7	23.6	2.7	3.8	30.3
390	Other manufactures	0.012	38	6.6	7.6	35.2	1.7	2.0	33.0

——— "Intraindustry Trade: Sources and Effects on Protection," *Journal of Political Economy* 95 (Dec. 1987), 1278–1291.

Nelson, F., and L. Olson, "Specification and Estimation of a Simultaneous Equation Model with Limited Dependent Variables," *International Economic Review* 19 (1978), 695–710.

Ray, Edward, "Tariff and Non-Tariff Barriers to Trade in the United States and Abroad," this REVIEW (May 1981a), 161–168.

——— "The Determinants of Tariffs and Nontariff Trade Restrictions in the U.S.," *Journal of Political Economy* 89 (Feb. 1981b), 105–121.

Ray, Edward, and Howard Marvel, "The Pattern of Protection in the Industrialized World," this REVIEW (1985), 452–458.

Rodrik, Dani, "What Does the Political Economy Literature on Trade Policy (Not) Tell Us that We Ought to Know?," NBER Working Paper 4870 (Sept. 1994).

Summers, Robert, and Alan Heston, "Penn World Tables (March 5): An Expanded Set of International Comparisons, 1950–1988," *Quarterly Journal of Economics* 106 (May 1991), 327–368.

Trefler, Daniel, "Trade Liberalization and the Theory of Endogenous Protection: An Econometric Study of US Import Policy," *Journal of Political Economy* 101:1 (1993), 138–160.

UNCTAD, *Handbook of Trade Control Measures of Developing Countries*, primary volume and supplement (1987).

——— *UNCTAD Micro TCM System* (1991).

We should note that outlying observations in tables 1 and 2 are the effect of particularly large protection by specific countries in certain industries. For example, despite otherwise moderate barriers, Venezuela protected its furniture industry in 1988 with an 85% tariff and 93% NTB coverage, accounting for the large standard deviations of Venezuela's protection found in table 1. Similarly, Sweden had zero or near-zero NTB coverage for most industries, but substantial protection (coverage ratios over 50%) in food, textiles, clothing, and footwear. Finally, the United Nations concordance we use to go from the detailed CCCN classification to the three-digit ISIC classification assigns no products to ISIC category 356, plastic products, but instead places these in various other categories.

V. Empirical Results

Equations (2) and (5) are estimated jointly using a simultaneous-equations tobit estimator, where the import equation (5) is the usual linear model, and the NTB equation (2) is censored at zero. The results are found in tables 3 and 4. As discussed above, all exogenous variables are used as instruments, with the political-economy variables excluded from the import equation, and the production share and distance measure excluded from the NTB equation. Also as noted before, factor endowments are used to instrument for the production share; the R^2 for these first-stage regressions is over 0.8 for all specifications.

In order to examine the degree to which industry- and country-specific factors determine protection, we estimate the models first with industry fixed effects, and then with both industry and country fixed effects. As a prelude to joint estimation of the simultaneous system, we briefly discuss the results of an analysis of variance (ANOVA) model for the two equations, in which we examine the extent to which industry and country fixed effects alone account for the variance of NTBs and imports.[3] For the NTB equation, the model with the two sets of fixed effects has an R^2 of 0.675, with the country effects alone accounting for 0.571. This suggests that a large part of the structure of NTBs can be best explained by simply looking at the average level of protection in each country, much as listed in table 1. In the import equation, on the other hand, the R^2 of the model with both country and industry effects is 0.566, with the industry effects accounting for 0.293, or just over half of the variance.

A. *Determinants of Nontariff Barriers*

Table 3 contains the results for estimation of the NTB equation (2). Columns 1 to 3 include only industry effects, whereas columns 4 to 6 include both industry and country fixed effects. Even with the inclusion of both sets of fixed effects, the results are in general well reconciled with political-economy theories of protection. As shown in the table, we present the variables in the following loose groupings: demand for protection, comparative advantage, political importance, and other trade influences. We discuss each group in turn.

The coefficient on the share of imports is significant and positive in all specifications except column 5, showing that industries "threatened" by import competition tend to receive more protection. The coefficient is even larger in magnitude in column 5 (which includes country effects and uses the labor share as the proxy for an industry's political importance), though it is far less precisely estimated.

The negative coefficients on the change in wage per worker indicate that declining industries receive more protection. Note, however, that this result is not robust across specifications: the coefficients are not statistically significant in the specifications which include country effects (columns 4–6), with the least significant result in column 5, where the share of workers replaces the share of value added. The coefficient on the share of industry output that is exported is negative in all specifications and statistically significant except in column 5. This accords with the idea that nations refrain from protecting industries for which exports are important out of fear that their trading partners will retaliate for any import restraints.

In columns 1 and 2, which include only industry fixed effects, the coefficient on wages is negative and significant, whereas the coefficient on labor productivity is positive and significant. This indicates that nations give less protection to industries with labor rents—industries in which wages are high after controlling for productivity. The significant negative coefficient on wages alone in column 3 similarly indicates that nations tend to provide more protection to low-wage ("weak") industries. We obtain a similar result (though somewhat less significant) if productivity alone is used instead of wages. The negative coefficient on wages somewhat allays our fears regarding the endogeneity of our right-hand-side variables, since we would usually expect wages to rise if they were affected by trade barriers. Note, however, that these results do not remain when we add country fixed effects in columns 4–6: in all three specifications, the coefficients become not significantly different from zero.

The coefficient on the industry share of value added is positive and significant in all specifications, indicating that large industries, which we interpret as politically important, also tend to receive more protection. In columns 2 and 5 we use the industry share of labor as the measure of political importance—this might better reflect the actual number of voters and thus the raw political importance of the sector rather than the economic importance indicated by the share of value added. The coefficient on the share of workers is positive and significant with just industry effects in column 2; but as with the import share above, when we add country effects, the coefficient on labor share becomes much larger in magnitude, but not statistically significant. Overall,

---

[3] Note that the ANOVA is a linear model and thus does not take into consideration the censoring of the NTM.

TABLE 3.—DETERMINANTS OF NONTARIFF BARRIERS
(DEPENDENT VARIABLE: 1 + NTB COVERAGE RATIO, 1030 OBSERVATIONS)

Independent Variables	(1)	(2)	(3)	(4)	(5)	(6)
		Industry			Industry and Country	
Demand for protection						
Imports versus domestic use	0.083 (4.14)	0.094 (4.33)	0.070 (3.86)	0.231 (3.07)	0.784 (0.93)	0.227 (2.95)
Change in wage per worker	−0.071 (−1.95)	−0.050 (−1.33)	−0.085 (−2.40)	−0.071 (−1.13)	−0.039 (−0.21)	−0.071 (−1.13)
Share of exports in gross output	−0.016 (−4.77)	−0.017 (−4.84)	−0.016 (−4.65)	−0.023 (−2.42)	−0.079 (−0.91)	−0.023 (−2.38)
Comparative advantage						
Wage per worker	−0.090 (−4.24)	−0.119 (−4.81)	−0.033 (−3.51)	0.012 (0.24)	−0.076 (−0.53)	0.032 (0.87)
Labor productivity (value added per worker)	0.066 (2.97)	0.103 (3.77)		0.020 (0.67)	0.424 (0.95)	
Political importance						
Share of value added	0.040 (2.66)		0.041 (2.75)	0.118 (2.62)		0.118 (2.62)
Share of workers		0.058 (3.29)			0.453 (0.89)	
Other trade influences						
1 + tariff rate	0.484 (8.81)	0.512 (8.82)	0.462 (8.66)	0.309 (3.51)	0.601 (1.26)	0.306 (3.47)
1 + black-market premium	0.554 (7.90)	0.579 (7.98)	0.538 (7.80)			
Import coefficient from single-equation estimate	−0.031 (−4.52)	−0.032 (−4.58)	−0.031 (−4.64)	0.002 (0.27)	−0.001 (−0.15)	0.002 (0.32)
Hausman test statistic (significance level)	36.523 (0.60)	37.621 (0.65)	36.051 (0.63)	9.434 (0.00)	0.881 (0.00)	8.733 (0.00)
Degrees of freedom for Hausman test	35	35	34	73	73	72

Notes: (1) Tobit equation in simultaneous-equations tobit estimation.
(2) *t*-statistics in parentheses, except for Hausman test.
(3) All variables in logs.
(4) A low significance level for the Hausman test means that the null hypothesis of no simultaneity bias cannot be rejected.

however, these results point to large industries receiving protection.

The positive and significant coefficients on tariffs and the black-market premium indicate that other trade measures and exchange rate controls are used in conjunction with NTBs; this matches the findings of Marvel and Ray. Note that we omit the black-market premium in columns 4–6, since this does not vary across industries within a country.

The bottom of table 3 provides the coefficient on imports from the corresponding single-equation Tobit estimates of equation (2), along with Hausman tests of the null hypothesis of no simultaneity bias for the NTB equation. We cannot reject this null at the 5% significance level for any of the six specifications—the *p*-values range from 0.40 to 0.35 with industry effects, and are uniformly one with both country and industry effects. On the other hand, comparing the two estimates of the coefficient on imports provides some evidence that the simultaneity of trade barriers and trade flows matters for the determinants of trade barriers. In the single-equation estimates, where the effect of NTBs on imports is neglected, we find either an insignificant or a negative effect of imports on NTBs; this no doubt reflects the usual import-reducing effects of NTBs. But this is strongly reversed in the simultaneous-equations estimates, in which we find a positive effect of import penetration on NTBs. In other words, our results indicate that it is not simply that industries with high import penetration receive protection, but rather that industries receive more protection to the extent that they have high import penetration after taking into account the level of NTBs.

B. *Determinants of Manufactured Imports*

Table 4 contains estimates of the specifications of the import equation (5); the columns correspond to those in table 3. As in the NTB equation, we drop distance and the black-market premium from columns 4–6, since these do not vary within each country.

The negative and highly significant coefficient on the output share in all specifications indicates that the monopo-

TABLE 4.—DETERMINANTS OF MANUFACTURED IMPORTS
(DEPENDENT VARIABLE: IMPORTS VERSUS DOMESTIC USE, 1030 OBSERVATIONS)

Independent Variables	(1)	(2)	(3)	(4)	(5)	(6)
	Industry			Industry and Country		
Output share	−0.394 (−13.00)	−0.410 (−13.65)	−0.389 (−12.72)	−0.439 (−10.19)	−0.550 (−10.43)	−0.442 (−9.88)
Distance	−1.053 (−8.30)	−1.081 (−8.57)	−1.032 (−8.07)			
1 + tariff rate	1.119 (2.02)	1.290 (2.34)	0.929 (1.61)	−0.954 (−2.29)	−0.021 (−0.05)	−0.895 (−2.04)
1 + NTB coverage ratio	−4.492 (−4.85)	−4.866 (−5.29)	−4.129 (−4.23)	0.796 (0.61)	−2.906 (−1.90)	0.567 (0.39)
1 + black-market premium	−0.202 (−0.31)	−0.031 (−0.05)	−0.414 (−0.61)			
Adjusted R^2	0.500	0.512	0.496	0.601	0.604	0.600
NTB coefficient from single-equation estimate	−1.150 (−5.96)	−1.148 (−6.00)	−1.146 (−5.95)	0.302 (1.26)	0.227 (0.94)	0.309 (1.28)
Hausman test statistic (significance level)	26.879 (0.28)	29.784 (0.42)	23.793 (0.15)	147.853 (0.99)	7452.117 (0.99)	5288.16 (0.99)
Degrees of freedom for Hausman test	32	32	32	69	69	69

Notes: (1) Linear equation in simultaneous-equations tobit estimation.
(2) t-statistics in parentheses, except for Hausman test.
(3) All variables in logs.
(4) A low significance level for the Hausman test means that the null hypothesis of no simultaneity bias cannot be rejected.

listic competition model seems to work well. Interestingly, estimation of the equivalent specifications without any fixed effects (results not shown) results in a substantially worse fit—the R^2 values are always less than 0.2. Adding industry fixed effects substantially improves the fit of the equation over the no-fixed-effects case. As shown in table 4, the adjusted R^2 values are about 0.5 with industry effects and just over 0.6 with both industry and country effects. The sizable contribution of the industry effects in explaining the import share matches the results of the ANOVA discussed above, and most likely indicates that the products of certain industries are simply more frequently traded than others, or that global trade in certain industries is comprised to a large degree of two-way shipments of differentiated products of the type best characterized by the monopolistic competition model. As expected, the coefficient on distance is always significant and negative; the coefficient on the black-market premium is similarly negative, but not statistically significant.

The specifications of the import equation in table 4 provide mixed evidence on the extent to which tariffs and NTBs reduce imports. When only industry fixed effects are included, NTBs have a strong negative effect on imports, and appear to be more substantial barriers to imports than tariffs, which have a statistically insignificant effect on imports. When country effects are added, however, the coefficients on tariffs and NTBs change in both sign and significance. In the specifications which use the industry share of value added (columns 4 and 6), the coefficient on tariffs becomes negative and significant, whereas that on NTBs is positive but not significant. When the labor share is used in column 5, however, the situation is nearly the opposite, with NTBs having a nearly significant effect on imports but tariffs almost no effect at all. In results not shown, we find that including both the labor share and the share of value added gives negative but insignificant coefficients on both NTBs and tariffs. Overall, our results seem to indicate that trade barriers in general hinder imports, but provide no conclusive evidence as to the relative impact of tariffs versus NTBs.

In contrast to our somewhat mixed results, Harrigan finds that "tariffs . . . were a more substantial barrier to trade in manufactures between developed countries than were non-tariff barriers." Of course, our specification is substantially different, not only in the number of countries, but also in that Harrigan estimates industry-specific regressions using bilateral trade flows, whereas we have data only on the tota imports and exports of each country. As mentioned before because we pool across industries, our results in table 4 should be interpreted as average values of the effects of trade barriers on imports across countries and industries. Evaluating at the average values of our data, the coefficients in columns 1–3 imply that if the global coverage of NTB across industries decreases by 10%, then the import share will rise by between 7.4% and 8.9%, whereas the results in column 5 imply an increase of 5.2% in the import share These results are quite close to the figure of around 5.0% tha Trefler (1993) obtains for the United States.

The Hausman tests at the bottom of table 4 show that w cannot reject the null hypothesis of no simultaneity bias i the import equation when we include only industry effects but strongly reject the null in the specifications with bot industry and country effects. Even though we cannot reject the null of no simultaneity in the specifications with only industry effects, allowing for the simultaneity of NTBs and imports in these specifications results in estimates of the effect of NTBs on imports which are two to three times larger in absolute value than the single-equation estimates which ignore this simultaneity. In the specifications with both industry and country fixed effects, the Hausman tests provide statistical evidence that this simultaneity is important, though our inconclusive results for the effect of NTBs on imports preclude a straightforward interpretation of differences in this coefficient between single-equation and simultaneous-equations estimates.

VI. Conclusion

Our results indicate that protection is not specific to particular countries and industries, but instead that the structure of NTBs across countries and industries can be explained by sectoral conditions. This is consistent with political-economy explanations of trade protection. Of course, we have not tested a specific model of protection, but rather examined some of its general determinants.

Also, we have data for only a single cross section of data on tariffs and NTBs, and are thus not able to look at the effects of changes in protection over time. This is clearly a concern for our identifying restrictions, which use industry conditions such as labor productivity and wages per worker to instrument for the level of NTBs. If protective measures are long standing, the causality might be the reverse; that is, the existence of barriers could influence industry conditions, rather than policy makers responding to industry-specific calls for protection. However, the significant negative coefficients we find on wages per worker and changes in wages per worker somewhat mitigate this concern, since we would expect wages to rise rather than fall in response to protection.

Further work is needed to examine the impact of variables such as demand and supply elasticities, which figure prominently in the theoretical literature but for which we were unable to obtain data. Also, our measures of an industry's political importance—industry shares of labor and value added—are probably far from ideal. For some countries, unionization data might be an important measure, whereas for others, the extent of ownership by the ruling party or family would no doubt be closely tied to the structure of trade protection. And as mentioned before, bilateral data on trade barriers and trade flows would allow for industry-by-industry estimation, whereas disaggregated data on factor endowments would allow us to compare the results of a Heckscher–Ohlin model with those from our monopolistic competition model of trade.

The limitations imposed by data availability notwithstanding, we obtain remarkably robust results that sectoral factors are important determinants of the structure of trade protection, even after taking into account industry- and country-specific fixed effects. Our results thus provide encouraging support for the burgeoning literature on the political economy of trade protection.

REFERENCES

Baldwin, Robert E., "The Political Economy of Protectionism," in Jagdish Bhagwati (ed.), *Import Competition and Response* (Chicago: University of Chicago Press, 1982), 263–286.

Barro, Robert J., and Jong-Wha Lee, "Data Set for a Panel of 138 Countries, 1960–198[illegible]," mimeo, Harvard University (Feb. 1993).

Bergstrand, Jeffrey, "The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence," this REVIEW 67 (1985), 474–481.

——— "The Generalized Gravity Equation, Monopolistic Competition, and the Factor-Proportion Theory in International Trade," this REVIEW 71:1 (1989), 143–153.

Bhagwati, Jagdish (ed.), *Import Competition and Response* (Chicago: University of Chicago Press, 1982).

Caves, Richard, "Economic Models of Political Choice: Canada's Tariff Structure," *Canadian Journal of Economics* 9 (1976), 278–300.

Dean, Judith, Seema Desai, and James Riedel, "Trade Policy Reform in Developing Countries since 1985: A Review of the Evidence," World Bank Discussion Paper 267 (May 1994).

Dick, Andrew, "Explaining Managed Trade: Non-Cooperative Politics or Cooperative Economics?" mimeo, UCLA (Jan. 1994).

Flam, Harry, and Elhanan Helpman, "Industrial Policy under Monopolistic Competition," *Journal of International Economics* 22 (1987), 79–102.

Gawande, Kishore, "Are U.S. Nontariff Barriers Retaliatory? An Application of Extreme Bounds Analysis in the Tobit Model," this REVIEW (November 1995), 677–688.

Grossman, Gene, and Elhanan Helpman, "Protection for Sale," *American Economic Review* 84:4 (Sept. 1994), 833–850.

Harkness, Jon, "Factor Abundance and Comparative Advantage," *American Economic Review* (Dec. 1978), 784–800.

Harrigan, James, "Factor Endowments & the International Location of Production: Micro-Econometric Evidence for the OECD, 1970–1985," *Journal of International Economics* 39 (1995), 123–141.

——— "OECD Imports and Trade Barriers in 1983," *Journal of International Economics* 35:1/2 (1993), 91–112.

Helleiner, G. K., "The Political Economy of Canada's Tariff Structure: An Alternative Model," *Canadian Journal of Economics* 10 (1977), 318–326.

Helpman, Elhanan, "Imperfect Competition and International Trade: Evidence from Fourteen Industrial Countries," *Journal of the Japanese and International Economies* 1 (1984), 62–81.

Helpman, Elhanan, and Paul Krugman, *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition and the International Economy* (Cambridge, MA: MIT Press, 1985).

Hufbauer, Gary C., and Howard F. Rosen, *Trade Policy for Troubled Industries* (Washington, DC: Institute for International Economics, 1986).

Laird, Sam, and Alexander Yeats, *Quantitative Methods for Trade Barrier Analysis* (New York: New York University Press, 1990).

Lawrence, Robert Z., "Does Japan Import Too Little: Closed Markets or Minds?," *Brookings Papers in Economic Activity,* no. 2 (1987), 517–554.

Leamer, Edward, "Cross Section Estimation of the Effects of Trade Barriers," in Robert Feenstra (ed.), *Empirical Methods for International Trade* (Cambridge, MA: MIT Press, 1988), chap. 3.

——— "Latin America as a Target of Trade Barriers Erected by the Major Developed Countries in 1983," *Journal of Development Economics* 32 (1990), 337–368.

Leamer, Edward, and Harry P. Bowen, "Cross-Section Tests of the Heckscher–Ohlin Theorem: Comment," *American Economic Review* (Dec. 1981), 1040–1043.

Magee, Stephen, William Brock, and Leslie Young, *Black Hole Tariffs and Endogenous Policy Theory.* (Cambridge: Cambridge University Press, 1989).

Marvel, Howard P., and Edward Ray, "The Kennedy Round: Evidence on the Regulation of International Trade in the United States," *American Economic Review* 73:1 (1983), 190–197.

Want to Know More?

To read further on this subject requires considerable selectivity, as the amount of literature on democracy has expanded even faster in recent years than the trend toward democracy itself. An excellent starting point is the new book by political scientists Juan J. Linz and Alfred Stepan, who address the difficult question of how societies can go beyond initial democratic openings to the actual consolidation of democracy: *Problems of Democratic Transition and Consolidation: Southern Europe, South America, and Post-Communist Europe* (Baltimore: Johns Hopkins University Press, 1996). Political scientist Philippe Schmitter seeks to go beyond "transitology" to "consolidology" in his "Transitology: The Science or the Art of Democratization?" in Joseph S. Tulchin, ed., *The Consolidation of Democracy in Latin America* (Boulder: Lynne Rienner, 1995). Larry Diamond, coeditor of the *Journal of Democracy*, offers a realistic, thorough exploration of whether the global democratic trend has come to an end in "Is the Third Wave Over?" *Journal of Democracy* 7, no. 3 (July 1996). The state of democracy in the world and the proper place of democracy promotion in the Clinton administration's foreign policy are examined in Thomas Carothers, "Democracy Without Illusions," *Foreign Affairs* 76, no. 1 (January/February 1997). Political scientists Adam Przeworski and Fernando Limongi present a systematic, eye-opening analysis of the relationship between economic development and democracy in their article "Modernization: Theories and Facts," *World Politics* 49, no. 2 (January 1997). Former French Foreign Ministry strategic affairs adviser Philippe Delmas, an accomplished French realist, casts a pox on the democracy house in his new book, *The Rosy Future of War* (New York: The Free Press, 1997), arguing that strong, stable states, and not democratic states, are the key to international peace and security.

For links to Web sites that offer still more useful and interesting information, visit FOREIGN POLICY's Web site at www.foreignpolicy.com.

Sense and Nonsense in the Globalization Debate

by Dani Rodrik

Globalization, Thomas Friedman of the *New York Times* has observed, is "the next great foreign policy debate." Yet as the debate expands, it gets more confusing. Is globalization a source of economic growth and prosperity, as most economists and many in the policy community believe? Or is it a threat to social stability and the natural environment, as a curious mix of interests ranging from labor advocates to environmentalists—and including the unlikely trio of Ross Perot, George Soros, and Sir James Goldsmith—argue? Has globalization advanced so far that national governments are virtually powerless to regulate their economies and use their policy tools to further social ends? Is the shift of manufacturing activities to low-wage countries undermining global purchasing power, thus creating a glut in goods ranging from autos to aircraft? Or is globalization no more than a

---

DANI RODRIK *is the Rafiq Hariri professor of international political economy at Harvard University. This article is based on the author's book,* Has Globalization Gone Too Far? *(Institute for International Economics, 1997).*

WHARTON REPROGRAPHICS

buzzword and its impact greatly exaggerated?

There are good reasons to be concerned about the quality of the globalization debate. What we are witnessing is more a dialogue of the deaf than a rational discussion. Those who favor international integration dismiss globalization's opponents as knee-jerk protectionists who do not understand the principle of comparative advantage and the complexities of trade laws and institutions. Globalization's critics, on the other hand, fault economists and trade specialists for their narrow, technocratic perspective. They argue that economists are too enamored with their fancy models and do not have a good handle on how the real world works. The result is that there is too much opponent bashing—and too little learning—on each side.

The ideological onslaught against the welfare state has made many governments unable to respond to the domestic needs of a more integrated economy.

Both sides have valid complaints. Much of the popular discussion about globalization's effect on American wages, to pick one important example, ignores the considerable research that economists have undertaken. A reasonably informed reader of the nation's leading op-ed pages could be excused for not realizing that a substantial volume of literature on the relationship between trade and inequality exists, much of which contradicts the simplistic view that Americans or Europeans owe their deteriorating fortunes to low-wage competition from abroad. The mainstream academic view actually is that increased trade with developing countries may account for at most 20 per cent of the reduction in the earnings of low-skilled American workers (relative to highly skilled workers) but not much more. One has to look elsewhere—to technological changes and deunionization, for example—to explain most of the increase in the wage gap between skilled and unskilled workers.

It is also true, however, that economists and proponents of trade have either neglected or pooh-poohed some of the broader complications associated with international economic integration. Consider the following questions: To what extent have capital mobility and the outsourcing of production increased the *substitutability* of domestic labor across national boundaries, thereby aggravating the economic insecurity confronting workers (in addition to exerting downward pressure on their wages)? Are the distributional implications of globalization—and certainly there are some—reconcilable with domestic concepts of distributive justice? Does trade with countries that have different norms and social institutions clash with and undermine long-standing domestic social bargains? To what extent does globalization undermine the ability of national governments to provide the public goods that their citizenries have come to expect, including social insurance against economic risks?

These are serious questions that underscore the potential of globally expanding markets to come into conflict with social stability, even as these markets provide benefits to exporters, investors, and consumers. Some of these questions have not yet been seriously scrutinized by economists. Others cannot be answered with economic and statistical analysis alone. But the full story of globalization cannot be told unless these broader issues are addressed as well.

THE LIMITS OF GLOBALIZATION

Even with the revolution in transportation and communication and the substantial progress made in trade liberalization over the last three decades, national economies remain remarkably isolated from each other. This isolation has a critical implication, which has been repeatedly emphasized by economist Paul Krugman: Most governments in the advanced industrial world are not nearly as shackled by economic globalization as is commonly believed. They retain substantial autonomy in regulating their economies, in designing their social policies, and in maintaining institutions that differ from those of their trading partners.

The supposition that domestic economies are now submerged in a seamless, unified world market is belied by various pieces of evidence. Take the case of North America. Trade between Canada and the United States is among the freest in the world and is only minimally hampered by transport and communications costs. Yet a study by Canadian economist John McCallum has documented that trade between a Canadian province and a U.S. state (that is, *international* trade) is on average 20 times smaller than trade between two Canadian provinces (that is, *intranational* trade). Clearly, the U.S. and Canadian markets remain substantially delinked from each other. And if this is true of U.S.–Canadian trade, it must be all the more true of other bilateral trade relationships.

The evidence on the mobility of physical capital also contradicts current thought. Popular discussions take it for granted that capital is now entirely free to cross national borders in its search for the highest returns. As economists Martin Feldstein and Charles Horioka have pointed out, if this were true, the level of investment that is undertaken in France would depend only on the profitability of investment in France, and it would have no relationship to the available savings in France. Actually, however, this turns out to be false. Increased savings in one country translate into increased investments in that country almost one for one. Despite substantial crossborder money flows, different rates of return among countries persist and are not equalized by capital moving to higher-return economies.

One can easily multiply the examples. U.S. portfolios tend to be remarkably concentrated in U.S. stocks. The prices of apparently identical goods differ widely from one country to another despite the fact that the goods can be traded. In reality, national economies retain a considerable degree of isolation from each other, and national policymakers enjoy more autonomy than is assumed by most recent writings on the erosion of national sovereignty.

The limited nature of globalization can perhaps be better appreciated by placing it into historical context. By many measures, the world economy was more integrated at the height of the gold standard in the late 19th century than it is now. In the United States and Europe, trade volumes peaked before World War I and then collapsed during the interwar years. Trade surged again after 1950, but neither Europe nor the United States is significantly more open today (gauging by ratios of trade to national income) than it was under the gold standard. Japan actually exports less of its total production today than it did during the interwar period.

GLOBALIZATION MATTERS

It would be a mistake to conclude from this evidence that globalization is irrelevant. Due to the increased importance of trade, the options

available to national policymakers have narrowed appreciably over the last three decades. The oft-mentioned imperative of maintaining "international competitiveness" now looms much larger and imparts a definite bias to policymaking.

Consider labor market practices. As France, Germany, and other countries have shown, it is still possible to maintain labor market policies that increase the cost of labor. But globalization is raising the overall social cost of exercising this option. European nations can afford to have generous minimum wages and benefit levels if they choose to pay the costs. But the stakes—the resulting unemployment levels—have been raised by the increased international mobility of firms.

The consequences are apparent everywhere. In Japan, large corporations have started to dismantle the postwar practice of providing lifetime employment, one of Japan's most distinctive social institutions. In France and Germany, unions have been fighting government attempts to cut pension benefits. In South Korea, labor unions have taken to the streets to protest the government's relaxation of firing restrictions. Developing countries in Latin America are competing with each other in liberalizing trade, deregulating their economies, and privatizing public enterprises.

Ask business executives or government officials why these changes are necessary, and you will hear the same mantra repeated over and over again: "We need to remain (or become) competitive in a global economy." As some of these changes appear to violate long-standing social bargains in many countries, the widespread populist reaction to globalization is perhaps understandable.

The anxieties generated by globalization must be seen in the context of the demands placed on national governments, which have expanded radically since the late 19th century. At the height of the gold standard, governments were not yet expected to perform social-welfare functions on a large scale. Ensuring adequate levels of employment, establishing social safety nets, providing medical and social insurance, and caring for the poor were not parts of the government agenda. Such demands multiplied during the period following the Second World War. Indeed, a key component of the implicit postwar social bargain in the advanced industrial countries has been the provision of social insurance and safety nets at home (unemployment compensation, severance payments, and adjustment assistance, for example) in exchange for the adoption of freer trade policies.

Forms of Social Insurance

All societies maintain provisions for social insurance to counter large drops in workers' and families' living standards, but that insurance takes different forms. In Europe and North America, income transfers paid out by the government are the predominant form of social insurance. Old-age pensions, unemployment compensation, disability insurance, and family support constitute the bulk of such transfers. The effect of these programs is twofold: They establish an income minimum for the citizenry regardless of employment status, and they reduce uncertainty regarding lifetime earnings for workers. In the United States, the Trade Adjustment Assistance (TAA) program targets workers who lose their jobs due to import competition. The TAA provides additional unemployment benefits, training subsidies, and relocation assistance. Comparatively few workers have benefited from TAA.

In Japan and other East Asian countries, income transfers are small compared with those in European nations. Many of the social insurance functions provided by the state in Western nations are in fact supplied by large enterprises in East Asia. These come in the form of lifetime-employment guarantees and employer-provided social services, ranging from housing and medical care to family support.

Many developing countries lack the administrative capacity to run income-transfer programs, and only a small share of the labor force is employed in the formal sector. In such countries, social insurance often takes yet another form: public-works programs and employment in the public sector, where jobs are typically more secure than in the private sector.

Government programs are not the only mechanism for reducing income risk. Private insurance, community support, and household transfers are also important. As markets spread and mobility increases, however, some of the informal mechanisms for alleviating income insecurity—such as community-based social services—will become harder to sustain.

This bargain is clearly eroding. Employers are less willing to provide the benefits of job security and stability, partly because of increased competition but also because their enhanced global mobility makes them less dependent on the goodwill of their local work force. Governments are less able to sustain social safety nets, because an important part of their tax base has become footloose because of the increased mobility of capital. Moreover, the ideological onslaught against the welfare state has paralyzed many governments and made them unable to respond to the domestic needs of a more integrated economy.

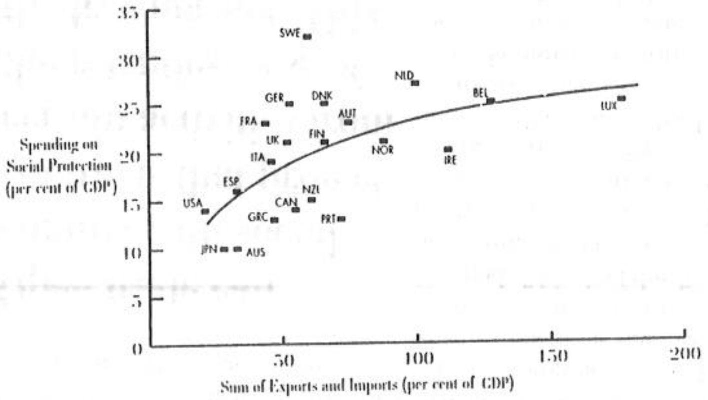
MORE TRADE, MORE GOVERNMENT

The postwar period has witnessed two apparently contradictory trends: the growth of trade and the growth of government. Prior to the Second World War, government expenditures averaged around 20 per cent of the gross domestic products (GDPs) of today's advanced industrialized countries. By the mid-1990s, that figure had more than doubled to 47 per cent. The increased role of government is particularly striking in countries like the United States (from 9 to 34 per cent), Sweden (from 10 to 69 per cent), and the Netherlands (from 19 to 54 per cent). The driving force behind the expansion of government during this period was the increase in social spending—and income transfers in particular.

It is not a coincidence that social spending increased alongside international trade. For example, the small, highly open European economies like Austria, the Netherlands, and Sweden have large governments in part as a result of their attempts to minimize the social impact of openness to the international economy. It is in the most open countries like Denmark, the Netherlands, and Sweden that spending on income transfers has expanded the most.

Indeed, there is a surprisingly strong association across countries between the degree of exposure to international trade and the importance of the government in the economy. The chart on page 26 shows the relationship between trade and spending on social protection (including unemployment insurance, pensions, and family benefits) in 21 countries for which the Organization for Economic Cooperation and Development (OECD) publishes crossnationally comparable data. The chart reveals an unmistakably positive correlation between a nation's openness to trade and the amount of its spending on social programs. At one end of the distribution we have the United States and Japan, which have the lowest trade shares in GDP and some of the lowest shares of spending on social protection. At the other end, Luxembourg, Belgium, and the Netherlands have economies with high degrees of openness and large income transfers. This relationship is not confined to OECD economies: Developing nations also exhibit this pattern. Furthermore, the extent to which imports and exports were important in a country's economy in the early 1960s provided a good predictor of how big its government would become in the ensuing three decades, regardless of how developed it was. All the available evidence points to the same, unavoidable conclusion: The social welfare state has been the flip side of the open economy.

RELATIONSHIP BETWEEN TRADE AND SPENDING ON SOCIAL PROTECTION (1980)



Source: Author

International economic integration thus poses a serious dilemma: Globalization increases the demand for social insurance while simultaneously constraining the ability of governments to respond effectively to that demand. Consequently, as globalization deepens, the social consensus required to keep domestic markets open to international trade erodes.

Since the early 1980s, tax rates on capital have tended to decrease in the leading industrial nations, while tax rates on labor have continued generally to increase. At the same time, social spending has stabilized in relation to national incomes. These outcomes reflect the tradeoffs facing governments in increasingly open economies: The demands for social programs are being balanced against the need to reduce the tax burden on capital, which has become more globally mobile.

By any standard, the postwar social bargain has served the world economy extremely well. Spurred by widespread trade liberalization, world trade has soared since the 1950s. This expansion did not cause major social dislocations and did not engender much opposition in the advanced industrial countries. Today, however, the process of international economic integration is taking place against a backdrop of retreating governments and diminished social obligations. Yet the need for social insurance for the vast majority of the population that lacks international mobility has not diminished. If anything, this need has grown.

The question, therefore, is how the tension between globalization and the pressure to mitigate risks can be eased. If the vital role that social insurance played in enabling the postwar expansion of trade is neglected and social safety nets are allowed to dwindle, the domestic consensus in favor of open markets will be eroded seriously, and protectionist pressures will soar.

THE GLOBAL TRADE IN SOCIAL VALUES

In the markets for goods, services, labor, and capital, international trade creates arbitrage—the possibility of buying (or producing) in one place at one price and selling at a higher price elsewhere. Prices thus tend to converge in the long run, this convergence being the source of the gains from trade. But trade exerts pressure toward another kind of arbitrage as well: arbitrage in national norms and social institutions. This form of arbitrage results, indirectly, as the costs of maintaining divergent social arrangements go up. As a consequence, open trade can conflict with long-standing social contracts that protect certain activities from the relentlessness of the free market. This is a key tension generated by globalization.

As the technology for manufactured goods becomes standardized and diffused internationally, nations with different sets of values, norms, institutions, and collective preferences begin to compete head on in markets for similar goods. In the traditional approach to trade policy, this trend is of no consequence: Differences in national practices and social institutions are, in effect, treated just like any other differences that determine a country's comparative advantage (such as endowments of physical capital or skilled labor).

In practice, however, trade becomes contentious when it unleashes forces that undermine the social norms implicit in domestic practices. For example, not all residents of advanced industrial countries are comfortable with the weakening of domestic institutions through the forces of trade, such as when child labor in Honduras replaces workers in South Carolina or when cuts in pension benefits in France are called for in response to the requirements of the Treaty on European Union. This sense of unease is one way of interpreting the demands for "fair trade." Much of the discussion surrounding the new issues in trade policy—e.g., labor standards, the environment, competition policy, and corruption—can be cast in this light of procedural fairness.

Open trade can conflict with social contracts that protect certain activities from the relentlessness of the free market. This is a key tension of globalization.

Trade usually redistributes income among industries, regions, and individuals. Therefore, a principled defense of free trade cannot be constructed without addressing the question of the fairness and legitimacy of the practices that generate these distributional "costs." How comparative advantage is created matters. Low-wage foreign competition arising from an abundance of workers is different from competition that is created by foreign labor practices that violate norms at home. Low wages that result from demography or history are very different from low wages that result from government repression of unions.

From this perspective it is easier to understand why many people are often ill at ease with the consequences of international economic integration. Automatically branding all concerned groups as self-interested protectionists does not help much. This perspective also prepares us not to expect broad popular support for trade when trade involves exchanges that clash with (and erode) prevailing domestic social arrangements.

Consider labor rules, for example. Since the 1930s, U.S. laws have recognized that restrictions on "free contract" are legitimate to counteract the effects of unequal bargaining power. Consequently, the employment relationship in the United States (and elsewhere) is subject to a multitude of restrictions, such as those that regulate working hours, workplace safety, labor/management negotiations, and so forth. Many of these restrictions have been put in place to redress the asymmetry in bargaining power that would otherwise disadvantage workers vis-à-vis employers.

Globalization upsets this balance by creating a different sort of asymmetry: Employers can move abroad, but employees cannot. There is no substantive difference between American workers being driven from their jobs by their fellow *domestic* workers who agree to work 12-hour days, earn less than the minimum wage, or be fired if they join a union—all of which are illegal under U.S. law—and their being similarly disadvantaged by *foreign* workers doing the same. If society is unwilling to accept the former, why should it countenance the latter? Globalization generates an inequality in bargaining power that 60 years of labor legislation in the United States has tried to prevent. It is in effect eroding a social understanding that has long been settled.

Whether they derive from labor standards, environmental policy, or corruption, differences in domestic practices and institutions have become matters of international controversy. That is indeed the common theme that runs the gamut of the new issues on the agenda of the World Trade Organization (WTO). Conflicts arise both when these differences create trade—as in the cases of child labor or lax environmental policies—and when they reduce it—as industrial practices in Japan are alleged to do. As the *New York Times* editorialized on July 11, 1996, in connection with the Kodak-Fuji dispute on access to the photographic film market in Japan, "the Kodak case asks the WTO, in effect, to pass judgment on the way Japan does business."

The notions of "fair trade" and "leveling the playing field" that lie behind the pressures for putting these new issues on the trade agenda have been ridiculed by economists. But once it is recognized that trade has implications for domestic norms and social arrangements and that its legitimacy rests in part on its compatibility with these, such notions are not so outlandish. These sentiments are ways of addressing the concerns to which trade gives rise. Free trade among countries with different domestic practices requires an acceptance of either an erosion of domestic structures or the need for some degree of harmonization or convergence.

If this is the appropriate context in which demands for "fair trade" or "leveling the playing field" must be understood, it should also be clear that policymakers often take too many liberties in justifying their actions along such lines. Most of the pricing policies that pass as "unfair trade" in U.S. antidumping proceedings, for example, are standard business practice in the United States and other countries. While there may not be a sharp dividing line between what is fair and unfair in international trade, one clear sign that pure protectionism is at the root of a trade dispute is the prevalence of practices within the domestic economy that are identical or similar to those being protested in the international arena. Fairness cannot be eliminated from thinking about trade policy; but neither can it be invoked to justify trade restrictions when the practice in question does not conflict with domestic norms as revealed by actual practice.

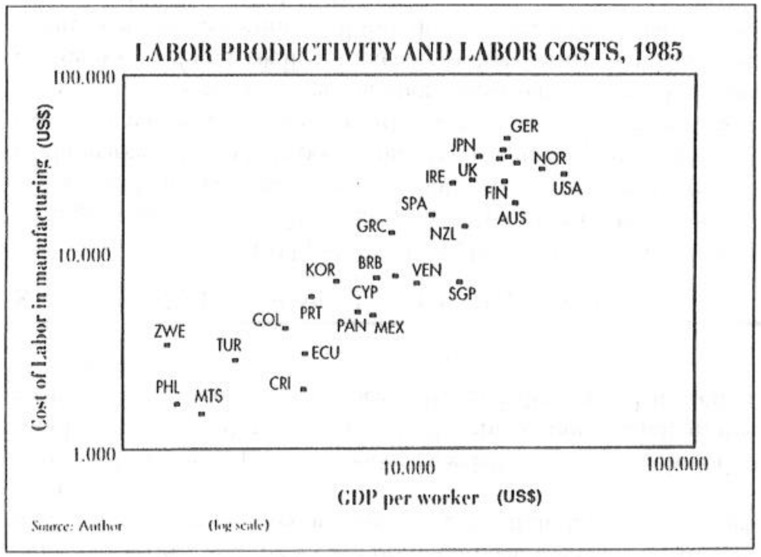
MISUNDERSTANDING TRADE

The tensions created by globalization are real. They are, however, considerably more subtle than the terminology that has come to dominate the debate. "Low-wage competition," "leveling the playing field," and "race to the bottom" are catchy phrases that often muddle the public's understanding of the real issues. A more nuanced debate and more imaginative solutions are badly needed.

A broader approach to this debate, one that takes into account some of the aspects discussed here, provides more credibility to the defenders of free trade in their attempts to clear up the misunderstandings that the opponents of trade often propagate. Journalist William Greider's recent book, *One World, Ready or Not—The Manic Logic of Global Capitalism*, illustrates the appeal that many of these misunderstandings retain in the minds of popular commentators on trade.

One of the main themes of this book—that the global expansion of markets is undermining social cohesion and is inexorably leading toward a major economic and political crisis—could be viewed as a more boldly expressed version of the potential danger that is highlighted above. Many of Greider's concerns—the consequences for low-skilled workers in the advanced industrial countries, the weakening of social safety nets, and the repression of political rights in some leading exporters like China and Indonesia—are indeed valid. However, the disregard for sound economic analysis and systematic empirical evidence that characterizes Greider's book makes it both a very unreliable guide to understanding what is taking place and a faulty manual for setting things right.

A popular fallacy perpetuated in works like Greider's is that low wages are the driving force behind today's global trade. If that were so, the world's most formidable exporters would be Bangladesh and a smattering of African countries. Some Mexican or Malaysian exporting plants may approach U.S. levels in labor productivity, while local wages fall far short. Yet what is true for a small number of plants does not extend to economies as a whole and therefore does not have much bearing on the bulk of world trade.


LABOR PRODUCTIVITY AND LABOR COSTS, 1985

Source: Author

The above chart shows the relationship between economy-wide labor productivity (GDP per worker) and labor costs in manufacturing in a wide range of countries. There is almost a one-to-one relationship between these two, indicating that wages are closely related to productivity. Low-wage economies are those in which levels of labor productivity are commensurately low. This tendency is of course no surprise to anyone with common sense. Yet much of the discourse on trade presumes a huge gap between wages and productivity in the developing country exporters.

Similarly, it is a mistake to attribute the U.S. trade deficit to the restrictive commercial policies of other countries—policies that Greider calls the "unbalanced behavior" of U.S. trading partners. How then can we explain the large U.S. deficit with Canada? If trade imbalances were determined by commercial policies, then India, as one of the world's most protectionist countries until recently, would have been running large trade surpluses.

Another misconception is that export-oriented industrialization has somehow failed to improve the livelihood of workers in East and Southeast Asia. Contrary to the impression one gets from listening to the opponents of globalization, life is significantly better for the vast majority of the former peasants who now toil in Malaysian or Chinese factories. Moreover, it is generally not the case that foreign-owned companies in developing countries provide working conditions that are inferior to those available elsewhere in the particular country; in fact, the reverse is more often true.

Perhaps the most baffling of the antiglobalization arguments is that trade and foreign investment are inexorably leading to excess capacity on a global scale. This is Greider's key argument and ultimately the main reason why he believes the system will self-destruct. Consider his discussion of Boeing's outsourcing of some of its components to the Xian Aircraft Company in China:

> When new production work was moved to Xian from places like the United States, the global system was, in effect, swapping highly paid industrial workers for very cheap ones. To put the point more crudely, Boeing was exchanging a $50,000 American machinist for a Chinese machinist who earned $600 or $700 a year. Which one could buy the world's goods? Thus, even though incomes and purchasing power were expanding robustly among the new consumers of China, the overall effect was an erosion of the world's potential purchasing power. If one multiplied the Xian example across many factories and industrial sectors, as well as other aspiring countries, one could begin to visualize why global consumption was unable to keep up with global production.

An economist would rightly point out that the argument makes little sense. The Chinese worker who earns only a tiny fraction of his American counterpart is likely to be commensurately less productive. Even if the Chinese worker's wages are repressed below actual productivity, the result is a transfer in purchasing power—to Boeing's shareholders and the Chinese employers—and not a diminution of global purchasing power. Perhaps Greider is thinking that Boeing's shareholders and the Chinese employers have a lower propensity to consume than the Chinese workers. If so, where is the evidence? Where is the global surplus in savings and the secular decline in real interest rates that we would surely have observed if income is going from low savers to high savers?

It may be unfair to pick on Greider, especially since some of his other conclusions are worth taking seriously. But the misunderstandings that his book displays are commonplace in the globalization debate and do not help to advance it.

SAFETY NETS, NOT TRADE BARRIERS

One need not be alarmed by globalization, but neither should one take a Panglossian view of it. Globalization greatly enhances the opportunities available to those who have the skills and mobility to flourish in world markets. It can help poor countries to escape poverty. It does not constrain national autonomy nearly as much as popular discussions assume. At the same time, globalization does exert downward pressure on the wages of underskilled workers in industrialized countries, exacerbate economic insecurity, call into question accepted social arrangements, and weaken social safety nets.

There are two dangers from complacency toward the social consequences of globalization. The first and more obvious one is the potential for a political backlash against trade. The candidacy of Patrick Buchanan in the 1996 Republican presidential primaries revealed that protectionism can be a rather easy sell at a time when broad segments of American society are experiencing anxieties related to globalization. The same can be said about the political influence of Vladimir Zhirinovsky in Russia or Jean-Marie Le Pen in France—influence that was achieved, at least in part, in response to the perceived effects of globalization. Economists may complain that protectionism is mere snake oil and argue that the ailments require altogether different medicine, but intellectual arguments will not win hearts and minds unless concrete solutions are offered. Trade protection, for all of its faults, has the benefit of concreteness.

Perhaps future Buchanans will ultimately be defeated, as Buchanan

Sons of Globalization?

800 GO PAT GO
Buchanan President
Buchanan President
GUN
WE LOVE PAT !!

Patrick Buchanan, Agence France-Presse/Corbis-Bettmann

Vladimir Zhirinovsky, Reuters/Jacky Naegelen/Archive Photos

VAS-Y JEAN-MARIE
LE PEN

Jean-Marie Le Pen, Corbis-Bettmann

himself was, by the public's common sense. Even so, a second and perhaps more serious danger remains: The accumulation of globalization's side effects could lead to a new set of class divisions—between those who prosper in the globalized economy and those who do not; between those who share its values and those who would rather not; and between those who can diversify away its risks and those who cannot. This is not a pleasing prospect even for individuals on the winning side of the globalization divide: The deepening of social fissures harms us all.

National policymakers must not retreat behind protectionist walls. Protectionism would be of limited help, and it would create its own social tensions. Policymakers ought instead to complement the external strategy of liberalization with an internal strategy of compensation, training, and social insurance for those groups who are most at risk.

In the United States, President Bill Clinton's education initiatives represent a move in the right direction. However, the August 1996 welfare reform act could weaken social safety nets precisely at a time when globalization calls for the opposite. In Europe, as well, the pruning of the welfare state may exacerbate the strains of globalization.

Contrary to widespread belief, maintaining adequate safety nets for those at the bottom of the income distribution would not break the bank. Currently, old-age insurance is the most expensive income-transfer item for the advanced industrial countries. A reorientation of public resources away from pensions and toward labor-market and antipoverty programs would be a more appropriate way to address the challenges of globalization. This shift could be achieved while reducing overall public spending. Broad segments of the population in the industrial countries are understandably nervous about changing basic social-welfare arrangements. Therefore, political leadership will be required to render such changes palatable to these groups.

At the global level, the challenge is twofold. On the one hand, a set of rules that encourages greater harmonization of social and industrial policies on a voluntary basis is needed. Such harmonization could reduce tensions that arise from differing national practices. At the same time, flexibility sufficient to allow selective disengagement from multilateral disciplines needs to be built into the rules that govern international trade.

Currently, the WTO Agreement on Safeguards allows member states to impose temporary trade restrictions following an increase in imports —but only under a stringent set of conditions. One could imagine expanding the scope of the agreement to include a broader range of circumstances, reflecting concerns over labor standards, the environment, and even ethical norms in the importing country. The purpose of such an expanded "escape clause" mechanism would be to allow countries—under well-specified contingencies and subject to multilaterally approved procedures—greater breathing room to fulfill domestic requirements that conflict with free trade. If this flexibility could be achieved in exchange for a tightening of rules on antidumping, which have a highly corrosive effect on the world trading system, the benefits could be substantial.

Globalization is not occurring in a vacuum: It is part of a broader trend we may call marketization. Receding government, deregulation, and the shrinking of social obligations are the domestic counterparts of the intertwining of national economies. Globalization could not have advanced this far without these complementary forces at work. The broader challenge for the 21st century is to engineer a new balance between the market and society—one that will continue to unleash the creative energies of private entrepreneurship without eroding the social bases of cooperation.

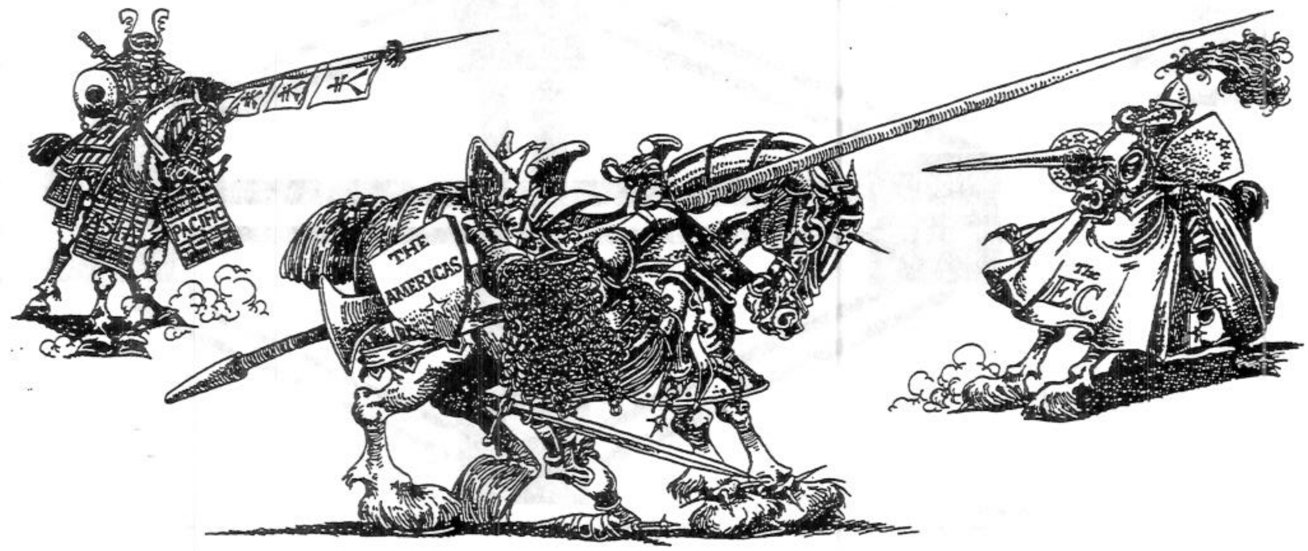
Want to Know More?

A classic work on the relationship between markets and social stability is Karl Polanyi, *The Great Transformation* (Boston, Massachusetts: Beacon Press, 1944). The argument is restated in the context of the postwar multilateral international economic order in John Ruggie, "At Home Abroad, Abroad at Home: International Liberalization and Domestic Stability in the New World Economy," *Millennium: Journal of International Studies*, no. 3 (1995). The close relationship between exposure to foreign trade and government size in the advanced industrial countries was first pointed out by David Cameron in "The Expansion of the Public Economy," *American Political Science Review* 72 (December 1978). A more recent study that is also more comprehensive in scope is Dani Rodrik, "Why Do More Open Economies Have Bigger Governments?" Working Paper no. 5537 (Cambridge, Massachusetts: National Bureau of Research, 1996).

An excellent historical perspective on the debates surrounding trade policy is provided in Jagdish Bhagwati, *Protectionism* (Cambridge, Massachusetts: MIT Press, 1988). An extended populist critique of globalization can be found in William Greider, *One World, Ready or Not—The Manic Logic of Global Capitalism* (New York: Simon & Schuster, 1997). Paul Krugman and Robert Lawrence, in "Trade, Jobs and Wages," *Scientific American* 270, no. 4 (April 1994), provide a succinct essay arguing that foreign competition is not responsible for the current ills of the U.S. economy. The recent academic debates on trade and wages are reviewed in William Cline, *Trade and Wage Inequality* (Washington, D.C.: Institute for International Economics, 1997, forthcoming) and Susan Collins, ed., *Trade and the American Worker* (Washington, D.C.: Brookings Institution, 1997, forthcoming). For links to Web sites related to globalization, see FOREIGN POLICY's home page at www.foreignpolicy.com.

WHARTON REPROGRAPHICS

WORLD TRADE



Jousting for advantage

The postwar drive to create a liberal system of global trade has run out of steam. Our economics editor, Clive Crook, weighs the implications

ECONOMIC policymakers from all over the world will gather in Washington, DC, on September 24th for the annual meetings of the International Monetary Fund and the World Bank. This year their routine discussions of monetary policy, exchange-rate co-ordination and international debt will be conducted against an unusually worrying background. Strife in the Middle East has once again cast its shadow over the world economy. The rise in oil prices, at least for a time, has increased the likelihood of slower economic growth and higher inflation.

There is precious little that governments in their role as economic policymakers can do about this except, teeth gritted, and with a firm hand on monetary policy, keep the inflationary momentum in check. But this does not mean that Washington's visitors need confine themselves to hand-wringing and statements of the obvious. Looking further ahead, they should recognise that the worsened outlook for the global economy has raised the stakes in another, and vital, area of international economic policy. It is all the more pressing that they address their minds to an issue which, for the moment, may seem less urgent, and where the economics is certainly more complicated—but on which they, if they choose, can make the crucial difference.

Over the past few years the international trading system—which, after all, has brought the industrialised economies to their present prosperity—has shown the first signs of unravelling. In the economic climate that lies ahead, the pressures on this system will be greatly intensified. If, over the next few years, liberal trade gives way to managed trade, if the global economy does splinter, as some fear, into separate trading blocks, and if the recent revival of protectionism accelerates, the economic losses could eventually put all three oil shocks to shame. The last great cycle of recession and protection reached its climax in the Great Depression of the 1930s.

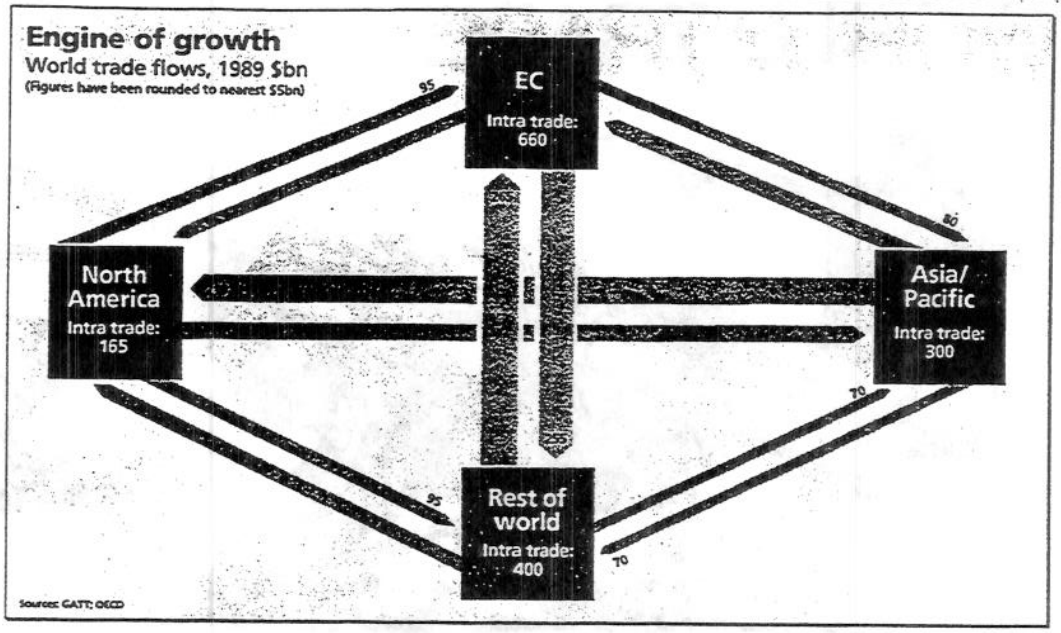
Finance ministers and central bankers have too often left the health of the international trading system in the hands of junior ministers and their civil servants. That has always been a big mistake, but never more so than now. Trade is the high ground of economic policy.

In Geneva for the past four years, the junior ministers and civil servants have been conducting a complicated series of negotiations on trade. These talks were launched in 1986 at a ministerial gathering in Punta del Este; hence they are known as the Uruguay round. They are held under the auspices of the General Agreement on Tariffs and Trade (GATT), the third and least illustrious of the Bretton Woods triplets. (What a trade minister is to a finance minister, the GATT is to the IMF and World Bank.) For most of the past four years the negotiators have agreed on little more than an extremely ambitious list of subjects to be discussed.

Time is running out. The talks are due to end in December. Governments must use next week's meetings in Washington to hurry things up. The future of the global trading system hangs on the outcome.

The new realists

Many politicians, especially in America, believe that the old approach to trade policy is already past saving. They agree with the congressmen who call the GATT the General Agreement to Talk and Talk—and Capitol Hill, admittedly, ought to know. Four years of discussion in Geneva have produced precious little evidence to challenge this view. The mood of disaffection goes beyond an understandable weariness with the GATT as an institution; it reflects a growing disenchantment with liberal trade. The new view, by no means confined to poli-



Engine of growth

World trade flows, 1989 $bn

(Figures have been rounded to nearest $5bn)

Source: GATT; OECD

ticians, is that the hands-off approach to trade is for suckers.

Nowadays many businessmen, while proclaiming their belief in free trade as a theoretical ideal, tell you at the same time that, as a matter of practical fact, free trade is a myth. Everybody is cheating. Trade agreements are meaningless. Competing means beating the successful exporters (especially the Japanese) at their own game—by promoting the domestic industries that matter, and by telling foreign governments to buy more imports, or else.

At the same time, the world economy shows signs of moving towards a trading system partitioned into three blocks: the Americas, the European Community and the Asia-Pacific region centred on Japan. The United States has signed its free-trade agreement with Canada and has started preliminary talks about similar deals with Mexico and Chile. Mr George Bush has broached the possibility of an American free-trade zone spanning the continent. The EC is busily engaged in Project 1992, which will create the world's largest single market for goods, services and capital. The economic changes in Eastern Europe, far from slowing the EC's progress down, seem to have speeded it up. If Japan and the East Asian economies find more of their goods shut out of Europe and America, an Asia-Pacific trade block may emerge by default.

Acronyms and all that

MFN	Most favoured nation
VER	Voluntary export restraint
VIE	Voluntary import expansion
USTR	United States trade representative
CAP	Common agricultural policy
AMS	Aggregate measure of support
MFA	Multi-Fibre Arrangement
TRIPS	Trade-related intellectual property rights
TRIMS	Trade-related investment measures
GATS	General agreement on trade in services

Against this background, the talks in Geneva have assumed a greater significance. For some, they are the system's last chance to prove itself. If the Uruguay round fails, the trade "realists" in Washington and elsewhere will claim to be vindicated. The GATT, and the multilateral, rules-based approach to trade that it stands for, may be not merely undermined but destroyed.

In a trivial sense, mind you, success in Geneva is assured. At the economic summit in Houston last July, America and its partners in the Group of Seven (G7) big industrial countries promised that they would not let the trade talks fail. They can keep this promise, because it is absurdly easy in trade negotiations to declare victory and retreat.

The summit itself was a case in point. In Houston the G7 leaders said they had agreed to a compromise on farm trade, one of the main sticking points in the round. It was soon clear that they had agreed only to fudge their continuing disagreements and thereby avoid an embarrassing public row. When governments make misleading claims about trade policy, they know that voters will rarely find them out. Deals in this area are so obscure and complicated that only people who are paid to understand them can be bothered to try—and of those who try, many fail.

Whatever the result of the talks, it is a good bet they will therefore be a "success"—maybe even a historic one. But the people who are already saying, like Mr Lester Thurow of the Massachusetts Institute of Technology, that "the GATT is dead," will hardly be impressed by an empty show of agreement. In that, at least, they would be right.

Impressive communiqués aside, governments do have goals for the Uruguay round. If they could secure it, they would all prefer the sort of success that delivers substantial, tangible benefits to their businesses back home: foreign markets opened here, playing fields levelled there. In other words, they hope for the same success that trade realists say will be won only when their governments play the game according to the new rules. The people these governments represent, who are consumers as well as producers, may not know it, but they have a big stake in yet another sort of success: a stable, liberal, non-discriminatory system of global trade.

The disjunction between the trade policies pursued by governments and the trade policies that are in the best interests of their citizens is the subject of this survey. For decades an international mechanism was in place to make it more likely that trade evolved in a way that was a success when viewed from either perspective. This mechanism was none other than the GATT. Lately, it seems, the mechanism has been breaking down. It is essential to ask whether the GATT has indeed failed and, if it has, why. Can it be repaired? What might replace it? Behind these questions lies an even bigger issue—one that economists and politicians have debated for centuries, but which seems to pose itself afresh every few decades. Is free trade a theoretical ideal of no practical use, or a workable policy for a troubled and imperfect world?

The ITO that never was

An unfinished history of the GATT

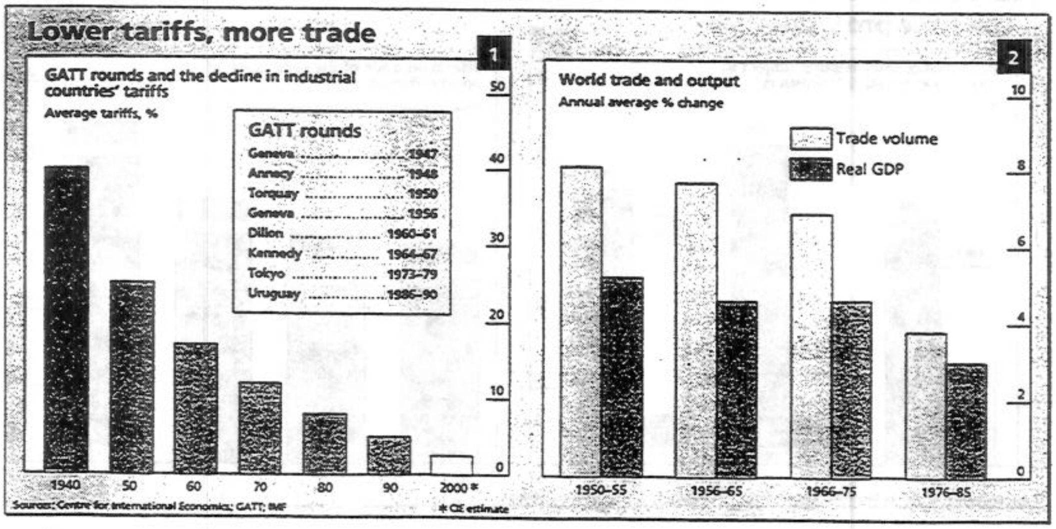
FOR most of its 43 years, the GATT has worked remarkably well. That is surprising, because the infant was four-fifths throttled at birth. The original designers, meeting at the Bretton Woods conference in 1944, envisaged a grand new institution, to be called the International Trade Organisation (ITO). It was to have been a full partner to the IMF and the World Bank, with departments covering employment policy, economic development, restrictive business practices and commodity agreements, as well as commercial policy (that is, tariffs and trade). But governments thought better of creating an economic organisation powerful enough to threaten their autonomy. The charter creating the ITO was never ratified. As a stop-gap, the governments settled on a compromise agreement about international trade.

Four decades later, this interim arrangement is still in place. A small secretariat—for the GATT is an institution as well as a treaty—monitors the trade policies of the members (23 of them at the beginning, now almost 100), serves the various committees that they set up, and helps to settle trade disputes.

Despite its awkward start, GATT did a fine job during its first 30 years. In a series of trade rounds, governments cut the average tariff on manufactured goods from 40% in 1947 to less than 10% by the mid-1970s. Since then the average tariff has fallen even further, to roughly 5% (chart 1). In addition, by the late 1950s, the industrial-country governments had established a modern system of international exchange and payments, with fully convertible currencies. As a result, tariff liberalisation spurred remarkably rapid growth in world trade. Between 1950 and 1975 the volume of trade expanded by as much as 500%, against an increase in global output of 220%.

The instruments of this spectacular change in the character of the world economy were the three guiding principles embodied in the articles of the GATT:

- **Reciprocity.** If one country lowers its tariffs against another's exports, it can expect the other country to lower its tariffs in return. This principle regards a market-opening measure as a concession, for which concessions in return are required. The idea is economic nonsense—as this survey will go on to argue—but it has proved terribly useful all the same. For the moment, note that the principle is exclusively concerned with bargains based on lower tariffs all round. It does not allow governments to threaten to raise their tariffs, even if such a threat might force other governments to lower theirs, thus levelling the playing field. Because this variety of reciprocity aims to match reductions in protection not levels of protection, economists call it "first-difference reciprocity".
- **Non-discrimination.** This principle says that countries should not grant one member or group of members preferential trade treatment over the others. It is known as the most-favoured-nation (MFN) rule. This does not mean, as the name might imply, that one or more countries are given an especially favoured status. It means the opposite, that every country is treated as favourably as the most favoured.
- **Transparency.** The GATT urges countries to replace non-tariff barriers (import quotas, for instance) with tariffs, and then to "bind" those tariffs (that is, promise not to raise them again). For any given amount of protection offered to domestic producers, non-tariff barriers do more economic harm than tariffs. Tariffs, especially when bound, foster a



Lower tariffs, more trade

1 GATT rounds and the decline in industrial countries' tariffs
Average tariffs, %

2 World trade and output
Annual average % change

Sources: Centre for International Economics; GATT; IMF

*CIE estimate

climate of greater certainty. And because their impact is easier to judge, they have the further advantage that they are much more amenable to negotiated reduction.

Reciprocity and non-discrimination, with a little help from transparency, are an exceptionally powerful double-act. Suppose a government is willing to lower its tariffs against imports from a certain country in order to gain access, thanks to reciprocity, to that country's market. If a deal can be struck, non-discrimination then requires that government to lower its tariffs on such imports regardless of which GATT member they come from. This, in turn, sparks another round of tariff-cutting, as the third parties are themselves obliged to reciprocate.

In this way, the benefits of bargains struck by the governments of big economies with lots of clout at the negotiating table are automatically extended to smaller countries, which might otherwise be left out. At the same time, the system makes sure that any initial spark of liberal intentions is fanned into a healthy flame. That was the theory and, in the areas of trade to which it was applied, that was how it worked.

The world's extraordinary economic progress since the second world war owes a great deal to the flame of international trade, and that in turn owes nearly everything to the GATT. Those members of America's Congress who nowadays deride it should remember what a power for economic good the GATT has been.

Something happened

Since the mid-1970s the volume of trade has continued to grow, but at a slower rate in relation to global output than before (chart 2 on previous page). That is a sign that something is amiss. The cause is not hard to find. The past 15 years have seen a proliferation of new sorts of trade protection.

One of these is the device that is cynically known as the "voluntary" export restraint (VER). Increasingly, if a powerful government is worried about the harm that imports are doing to its producers, it does not seek a remedy through the GATT. Instead, especially if it happens to be the American government, it requires another government, on pain of retaliation, to restrict its country's exports of the good in question.

In the late 1970s and 1980s the use of VERs spread from textiles and clothing to steel, cars, shoes, machinery, consumer electronics and more. The GATT's secretariat has counted nearly 300 VERs and similar arrangements. Most protect the American and European markets; 50 affect exports from Japan and another 35 exports from South Korea.

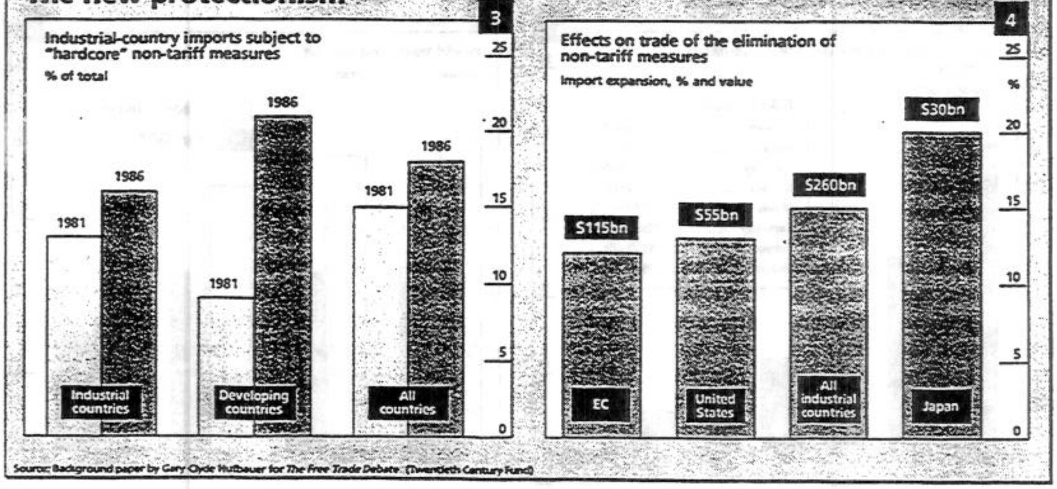
Because VERs and other so-called "hard-core" non-tariff measures are difficult to monitor (that is the point of them), there are no reliable estimates of their impact on trade. The World Bank reckons that the coverage of "hard-core" measures increased from 15% of the industrial countries' imports in 1981 to 18% by 1986. The Institute for International Economics in Washington, DC, guesses that if the industrial economies eliminated their non-tariff barriers, their exports would rise by nearly $200 billion a year and their imports by more than $250 billion (chart 3); the worldwide elimination of such measures might increase world trade by $330 billion a year (chart 4).

In the recent use of the VER, America has led the way, but other industrial countries have happily followed. The countries selflessly volunteering to restrict their exports include Japan, the newly industrialised East Asian economies, poorer countries such as Brazil, Mexico and Poland, and desperately poor ones such as Bangladesh. Members of the European Community, themselves devotees of the VER as an instrument of trade policy, are reluctant volunteers in one case: the Community has agreed to restrict its exports of steel to the United States.

So much for reducing trade. America also has ways of increasing it. The crowbar in America's toolbox is the notorious Section 301 of its trade law. This gives the president broad authority to retaliate against foreign trade practices that unfairly discourage American exports. America decides what is unfair; the law is vague on the point. The procedure can be started either by a private party's petition or at the request of the United States Trade Representative (USTR).

Between 1974, when the Trade Act became law, and 1985, 27 cases involving unfair barriers were brought. Since 1986, however, the law has been used more aggressively. In two years 11 cases were brought involving large amounts of trade. By 1988 $4 billion-worth of trade was subject to Section-301

The new protectionism



Source: Background paper by Gary Clyde Hufbauer for *The Free Trade Debate* (Twentieth Century Fund) action. If a VER is the remedy for excessive imports, then the apt reply to a Section-301 complaint would seem to be a VIE (voluntary import expansion). In its greater use of Section 301, America's trade policy has broadened in the 1980s from a purely defensive stance (concern about high imports) to include a more aggressive element (concern about low exports).

The Omnibus Trade and Competitiveness Act of 1988 made the old Section 301 more potent and added a new "Super 301". The difference between the two is that Section 301 deals with disputes about specific goods, whereas Super 301 is used to accuse countries of a broad range of unfair trade practices. Countries "named" under Super 301 must reach agreement with the USTR within 12 to 18 months or face retaliation. Japan, India and Brazil were named in 1989, though Japan and Brazil were later removed. Eight countries were put on a "priority watch list" because of their allegedly inadequate protection of intellectual property, and another 17 were told that they were being kept in mind for similar treatment.

Wielding this big stick, America negotiated its recent Structural Impediments Initiative with Japan. This extended the idea of unfair trade practices into previously unexplored territory. The negotiations covered Japan's retail-distribution system, its public spending on roads and infrastructure, households' saving habits, employment legislation and so on, the list extending to more than 200 issues.

Some economists complain that all this is merely "procedural protection"—using trade policies that are supposed to offer remedies for the abuses of other governments as a way of granting special treatment to your own producers. At any rate, a lot of what America has been up to is illegal under the GATT, or at least threatens to be so if carried through. Super 301 negotiations will probably bypass GATT procedures. Remedies are likely to violate all three of the principles described above: reciprocity, because America is not offering to lower its own trade barriers; non-discrimination, because America is implicitly threatening to block imports from one country; transparency, because the outcome is likely to involve some non-tariff intervention in trade flows.

America is by no means the only exponent of procedural protection—and not all the methods are inconsistent with the letter of the GATT. Article 19 of the GATT allows countries to take "safeguard" (or "escape-clause") actions to protect an industry that is being damaged by a sudden surge of imports. Such action must be temporary, non-discriminatory and transparent (tariffs, not quotas). All of which is jolly inconvenient.

Far more popular, therefore, are the rules on dumping and countervailing duties. Governments put anti-dumping duties on imports that are being sold in their market, they claim, at prices that are too low in relation either to costs or to prices in the exporter's home market. Similarly, countervailing duties can be imposed on imports from suppliers that are deemed to have been unfairly subsidised by their governments. In their initial effect, unlike Section 301 and Super 301, these rules are intended simply to reduce trade; they are not concerned with opening foreign markets.

Though anti-dumping and countervailing-duty measures attract far less popular attention than Section 301 and Super 301, they are much more often used. Between 1979 and 1988 the anti-dumping procedure accounted for 77% of all trade actions monitored by the GATT, and countervailing-duty cases another 18%. Over this period America brought 427 anti-dumping cases, 371 countervailing-duty cases, 44 Section-301 cases and 36 "others"; the European Community brought 406 anti-dumping cases, 13 countervailing-duty cases and 39 "others".

The GATT provides for anti-dumping and countervailing-duty remedies, even though these clash with the non-discrimination principle. They are a legal way to single out another member for higher-than-MFN tariffs. (The European Community makes no bones about this. It has long argued for a much greater relaxation of the non-discrimination principle.) On average, anti-dumping duties are four times higher than the corresponding MFN tariffs and one MFN tariff is replaced by four different rates of duty.

Many actions lead to price-fixing or quota arrangements. Two-thirds of the VERs imposed by America and the European Community between 1980 and 1987 were the outcome of anti-dumping actions. The mere threat of an action may have a big effect on trade. As Mr Patrick Messerlin, an economist at the World Bank, puts it: "Entire industrial sectors—steel and chemicals in the United States; steel, chemicals and electronics in the EC—are increasingly moulded by this instrument." All in all, it is difficult to believe that trade would not have grown a lot faster than it did during the 1980s if the accelerating slide towards procedural protection had never happened.

Compare and contrast

In 1990 many Americans agree with Mr Thurow that the GATT is dead, and favour "managed trade" as the only way to meet the economic challenge of the next decade. At the same time the countries of Eastern Europe are fleeing COMECON—managed trade in a fully worked-out form—and asking to join the largely open international trading system that the GATT has built. Odd, that.

The developing countries pose the same puzzle. As the industrial countries' enthusiasm for liberal trade cooled in 1980s, many of the world's poor countries began to see the advantages. This was quite a change.

From the outset, the third world enjoyed its "special and differential status" in the GATT. Article 18 of the agreement gave developing countries blanket permission to use trade restrictions to "protect" their balance of payments. Like the principle of reciprocity, this idea is an economic fallacy. Trade restrictions may succeed in reducing imports, but since they are just as likely to reduce exports, they have no reliable effect on the balance of trade. Unlike reciprocity, Article 18 has no mitigating rationale: it is the unacceptable face of the mercantilism that is built into the GATT. For years the poor countries eagerly exploited this permission to stay poor. As second-class citizens, spared the full, onerous duties of GATT membership, they meanwhile lacked the bargaining power to challenge the trade barriers erected against them.

The trade reformers of Latin America and Africa still have a long way to go, but a real change in thinking does seem to be under way. The third world's trade policies are still highly protectionist, it must be stressed. Equally, despite its recent change of direction, America remains one of the most open economies in the world. But it is striking that the trade-policy gulf between the richest economy in the world and the struggling economies to its south is narrowing fast from both sides.

What makes this even more striking is the central role played in both cases by the need to cut current-account deficits. America's deficit increased sharply in the early 1980s, as the dollar was driven up by high American interest rates; Latin America's current-account deficits were more deep-seated, reflecting years of over-borrowing and under-investment during the 1970s. America's policymakers wanted to reduce their country's external deficit partly to avoid a dangerous build-up of foreign debt, partly because they saw it as a sign of the pressure on domestic producers, and partly because it was an embarrassment; Latin America's governments were compelled to run current-account surpluses to service their debts. America turned somewhat protectionist; Mexico, Chile, Venezuela and others (including, more recently, Brazil) opened their markets to foreign competition.

Note too that the industrial countries' drift towards protection has happened alongside a marked shift in the opposite direction in their own domestic economic policies, from interventionism towards greater reliance on market forces. The idea of managed trade gained currency in the age of Ronald Reagan and Margaret Thatcher. It seems that over the past ten years economic liberalism has won the argument everywhere—except in the industrial countries' approach to trade.

Why? During the 30 years when the GATT was doing its job so splendidly, American leadership played the critical role. The United States was unchallenged as the world's foremost economic power. Its businesses had everything to gain if foreign markets could be opened to its exports, and little to lose, it seemed, if America's markets were opened to foreigners in return. All that has changed. The American economy is shrinking in relative terms, threatened from across the Pacific by Japan and from across the Atlantic by a resurgent Europe. Its economic supremacy, quantitative and qualitative, can no longer be taken for granted. America lacks the confidence to champion the cause of free trade; its government has lost the will to block demands for protection.

This, perhaps, is the best answer to the puzzle. But as Maynard Keynes said:

> The ideas of economists and political philosophers, both when they are right and when they are wrong, are more powerful than is commonly understood . . . I am sure that the power of vested interests is vastly exaggerated compared with the gradual encroachment of ideas.

In the 1980s a new body of economic theory began to challenge the classical arguments for liberal trade. Politicians who had never opened an economics textbook read about it on the editorial pages of their newspapers, and embraced it gratefully. Strategic trade theory did not "encroach". It arrived with a bang, an idea for its time.

The economics of free trade

How to make comparative advantage work for you

THE leading academic authority on the new trade theory is Mr Paul Krugman, of the Massachusetts Institute of Technology. He begins one of his articles on the subject* by saying that "the case for free trade is currently more in doubt than at any time since the 1817 publication of Ricardo's Principles of Political Economy." If this is true, it is partly because the challenge that the new theory poses to the classical case for free trade has been badly misunderstood. To see why, it is necessary to recall what the classical case for free trade actually says—because that theory, too, has been misunderstood.

These days politicians all over the world declare themselves in favour of free trade. When it comes to voting for it, they are not so sure. The reason is not just the pressure of special-interest politics. It is also that most people have imbibed the prejudice that free trade is a good thing, without imbibing the economics that ought to lie behind it. What this prejudice says, in fact, is that free trade is a good thing only if everybody else joins in; one-sided, or unilateral, free trade is a mug's game. The classical case for free trade argues exactly the opposite: free trade is good for a country even if other countries do not return the favour.

Writing 40 years before Ricardo, Adam Smith had already had a lot to say about the gains from trade. He saw it as, among other things, a way of promoting efficiency, both because it fostered competition and because it provided opportunites to specialise and gain economies of scale. Specialisation was a matter of absolute advantage: trade allows countries to produce what they are best at, and buy in the rest.

This view begged a question: what if Britain, say, is bad at making everything? Does this not mean that trade would drive all its producers out of business? David Ricardo answered the question by formulating the principle of comparative advantage. This is perhaps the single most powerful idea in economics.

Help yourself to wine and cheese

Suppose there are two countries, Utopia and Flatland, and that these countries use labour to produce just two goods, wine and cheese. In Utopia it takes one hour of labour to make a pound of cheese and two hours of labour to make a gallon of wine. In Flatland it takes six hours to make a pound of cheese and three hours to make a gallon of wine. Note that Utopia is more productive than Flatland in both goods; it has an absolute advantage in wine

---

*"Is Free Trade Passé?" *Journal of Economic Perspectives*, Vol 1, No 2, 1987

THE UTOPIAN FLATLANDISH FREE TRADE
WINE & CHEESE

and cheese. But its greater advantage, its comparative advantage, is in cheese. This will determine what happens when the two countries trade.

The precise outcome will depend on the pattern of demand, and hence on the price of each good in terms of the other once trade begins. Assume that a pound of cheese trades for a gallon of wine. This is for simplicity's sake; the argument does not turn on the price chosen. In Utopia, which is better at making both goods, an hour of labour can make either a pound of cheese or half a gallon of wine. But since a pound of cheese can be traded for a gallon of wine, it makes sense for Utopia to specialise in producing cheese, and then trade some of its cheese for wine. In this way it can consume as much cheese as before and twice as much wine, or some combination of more wine and cheese.

Flatland is less efficient than Utopia at making both goods. But in Flatland too it pays to specialise. An hour of its labour can make one-sixth of a pound of cheese or one-third of a gallon of wine (which is worth one-third of a pound of cheese in the international market). So Flatland specialises in the production of wine, and trades some of its wine for cheese. Trade means that it can consume as much wine as before and twice as much cheese, or some combination of more of both.

From the assumptions this example has already made, it is possible to deduce the market-determined wage that will be paid in each country. The hourly wage in Utopia will be a pound of cheese (equal to a gallon of wine) and the wage in Flatland will be one-third of a gallon of wine (equal to one-third of a pound of cheese). In other words, the wage in Utopia will be three times the wage in Flatland. This reflects the relative price of the two goods, together with the fact that Utopia is six times more productive in cheese and 1½ times more productive in wine.

Here is another way to think of it. Utopia is more efficient overall, so its wages are higher. This gives Flatland an opportunity to produce at lower cost, provided it specialises in wine, the good in which it has a comparative advantage (that is, a smaller absolute disadvantage). Utopia can make a gallon of wine with two hours of labour, whereas Flatland needs three—but because its wages are only one-third of Utopia's, Flatland can produce wine more cheaply. (Despite the fact that its wages are lower, Flatland cannot produce cheese more cheaply.)

Trade does not equalise incomes when productivity differs across countries, it just makes all sides better off than they would otherwise be. Moreover, trade always uncovers opportunities of this kind. Repeat, nothing central to the argument rests on the particular assumptions about price and productivity that have been used in this example.

The example has been borrowed, by the way, from an excellent textbook, "International Economics", written by Messrs Paul Krugman (the very same) and Maurice Obstfeld. You will find it on page 22—the principle of comparative advantage is not exactly advanced material. In parts of the real world, though, the free-trade debate still seems to be struggling with page xi (List of Contents). Messrs Krugman and Obstfeld quote with amusement an article from the *Wall Street Journal* ("The Coming Overthrow of Free Trade") which observed, "Many small countries have no comparative advantage in anything." One can imagine the writer prefacing that jewel of economic illiteracy with "Of course, we all believe in free trade, but . . ."

Please be realistic

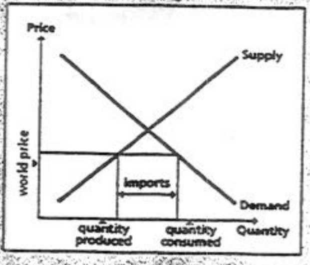
Many refinements and extensions of this simplest possible trade model (two countries, two goods, one factor of production) are needed to paint a lifelike

Hours required to make:

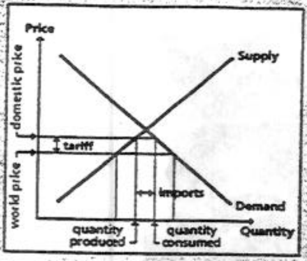
	1lb cheese	1 gallon of wine
In Utopia	1	2
In Flatland	6	3

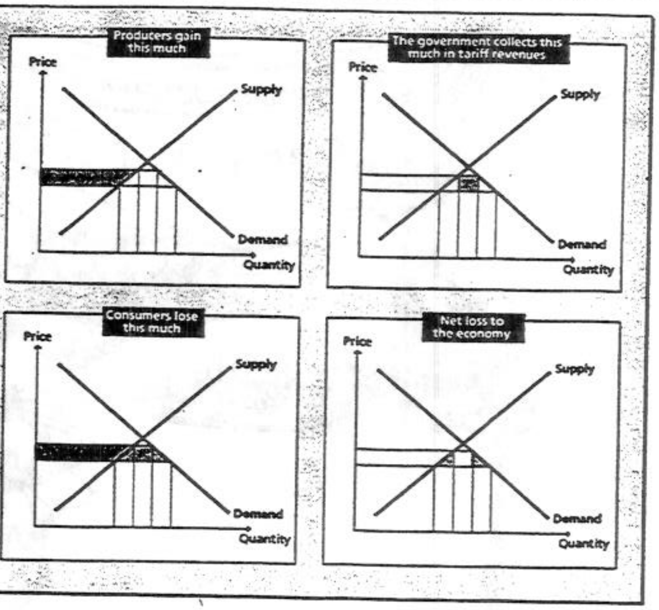
The geometry of free trade

Free trade
The amounts demanded and supplied are decided by the world price. If, at that price, demand exceeds domestic supply, imports bridge the gap.



Protection
A tariff raises the domestic price above the world price. Domestic supply rises and demand falls, reducing imports from both sides.





picture of the world. This picture has to include the fact that most countries are generalists, not narrow specialists. They produce, export and import a wide range of goods. The Ricardian theory was duly refined and extended. In its modern, sophisticated form it retains what is essential in Utopia and Flatland: (a) if trade is unimpeded, it will be driven by comparative advantage; (b) free trade makes countries better off. Note that these findings have nothing to do with reciprocity or "fair" trade.

In more modern guise, the classical case for unreciprocated free trade is summarised in the supply-and-demand diagrams in the box above. A tariff will increase a country's production of the protected good and raise the price paid to producers, who therefore gain. The government collects some extra revenues, too. But consumers buy less of the good than before and have to pay more for it, so they lose. Consumers lose more than the combined gain of producers and the government. Therefore stick with free trade.

Incidentally, an import quota is analysed in exactly the same way. Like a tariff, a quota raises prices in the domestic market. The only difference is that instead of collecting that handy block of tariff revenues, a government imposing a quota lets this money be gathered by the wicked foreigners to whom the quota applies. That is why the VER, a quota of sorts, is a particularly brainless way to protect.

Most modern classicists, in fact, would say that the gains from trade are certain to be a lot bigger than the two triangles in the last diagram. Protection takes the competitive pressure off producers, and therefore fosters inefficiency. The triangles ignore this crucial effect. Also, if practised on a large scale (as in Latin America in the 1960s and 1970s), protection denies the economy access to new techniques: imports often supply both the spur and the means to innovate.

Exceptions ancient and modern

The developing countries' bitter experience with autarkic plans of "import substitution", it might seem, should have settled the argument. Instead, the debate has flared up anew. However, trade-theory revisionists such as Mr Krugman (once he gets to page 261) were not the first to see difficulties with the case for free trade. Economists in the classical tradition have long allowed for qualifications to the idea that free trade is always and under all circumstances best.

One such qualification is the possibility of an "optimal tariff". The idea is simple. If a country is large enough, it may be able to depress the price of its imports on the world market by demanding less. In this way, it can use a tariff to improve its terms of trade (raise its export prices relative to its import prices). Conceivably, this could improve its welfare by more than the triangles and the other costs of protection reduce it. The optimal tariff would be the tariff that maximised this gain.

Another exception is the case for infant-industry protection. This idea (like many others) goes back to Adam Smith. It is often argued that new industries need to be sheltered when young. Without such help, enterprises that might be profitable in due course will fail to establish themselves. The infant-industry argument rests on the view that the market fails in some way; otherwise a new business would be willing to incur initial losses in the expectation of making big profits later.

Short-sighted financial markets might be one such market failure: they might make it impossible for a new business to finance its initial investment. Another sort of market failure happens when "externalities" are involved. A new company may incur costs in starting up that benefit others following it into the industry (for instance, the costs of adapting foreign technology for local use). The first company into the business is not fully rewarded because it cannot keep the benefits all to itself. As a result, too few new industries become established.

All these cases have long been recognised and analysed in the literature. Yet the case for free trade remained almost universally accepted by economists—no other proposition could command such support from that argumentative profession. The qualifications to the case for free trade were regarded as interesting but unimportant. The economic risks of trying to act on the exceptions were seen as out of all proportion to the likely costs. These included trade-policy retaliation by other countries (even though such retaliation would, in the classical view, be ill-advised), and the danger that protection for infants would be captured by vested interests and extended through adolescence, adulthood and premature senility. Behind it all was the unbudgeable view that an open international trading system was worth preserving.

The economics of managed trade

New theories claim to undermine the case for free trade. Do they?

IN THE main, the new theories broaden the old exceptions, to the point where the benefits to be gained from an interventionist trade policy might be worth the costs. The essential ingredient is the idea of imperfect competition. Many categories of trade, the argument goes, are dominated by a smallish number of big firms—big enough for each firm to have an influence on prices in the market. The classical case for free trade has its intellectual roots in the notion that firms are too small to move prices on their own. Things look different if you relax that assumption. The new trade theories fall under two main headings:

- **Strategic trade policy.** Imperfect competition means that firms may be able to earn economic "rents"—returns in excess of the economic costs (including a return on capital) of the resources used. Under perfect competition, such rents are competed away. The idea of strategic trade policy, which was first explored by Mr James Brander and Ms Barbara Spencer of the University of British Columbia, is that with a little bit of trade intervention, a government might secure a bigger share of these rents for its country.

The standard example is the aircraft industry—a good choice because start-up costs there are extraordinarily high (which keeps the number of possible producers down) and because national prestige is at stake. Suppose two companies, call them Boeing and Airbus, are choosing whether to build a new aircraft. If either goes ahead alone, it will make profits of 100. If both go ahead, they will each make a loss of 5.

Now suppose that Europe's governments agree to subsidise Airbus to the tune of 10. Airbus is sure to make a profit of 5 even if Boeing goes ahead too. But the point is this: Boeing will not go ahead. The subsidy makes it certain that Airbus will build its new aircraft, and certain therefore that Boeing will make a loss if it joins in. The result is that Airbus makes a profit of 110, and can repay its subsidy with ease. For the sake of argument, assume that each company produces entirely for export. Then Europe's subsidy of 10 has secured a transfer of rents worth 100 from America to Europe.

Consider another example. Imagine an industry with just two companies, one American and one Japanese. In this industry, learning-by-doing reduces unit costs indefinitely as output expands. Japan closes its domestic market to the American firm while America keeps its own market open. As a result, the Japanese firm can increase its output, thereby lowering its unit costs. In due course this competitive edge allows it to drive the American firm out of business. The rents that the American firm had extracted from American consumers now pass to the Japanese.

- **Externalities revisited.** This argument is much less novel than the idea of a strategic struggle over rents, because the classical case for free trade discusses externalities at length. The difference, argues Mr Krugman, is that theories based on perfect competition could not do justice to the most likely source of externality, whereas theories based on imperfect competition can.

The externality in question is the inability of firms to capture the fruits of expensive innovation. (Theories of perfect competition neglect this because large fixed costs are involved in R&D, and large fixed costs keep the number of competitors in an industry small.) This argument, a variant on the traditional infant-industry case, provides intellectual backing for a demand that is often heard in America and Europe, namely that governments should protect or otherwise support their high-tech industries.

Since such industries are often said to be "strategic", the jargon here has got itself into a twist. The "strategic" in strategic trade policy refers to a beggar-thy-neighbour struggle over rents. The externalities argument for supporting high-tech industries does not rely on this chance to grab rents (though for many of the industries concerned that case could doubtless be made as well); in principle, the world as a whole may benefit if a government promotes R&D that had previously been blocked by the externality. Whereas strategic trade policy is a zero-sum game, intervention to deal with externalities is a positive-sum game. In the popular debate, these two quite different rationales for intervention are cheerfully lumped together.

Caveat emptor

Some of the most effective attacks on using the new trade theory as a guide to policy have come from the trade-theory revisionists themselves. Mr Krugman and the others are well aware that their views have been hijacked by uncomprehending politicians, who have used them as a cloak of respectability for their protectionist instincts.

In a recent article, Mr Krugman wrote: "Most challenges to the *laissez-faire* view do not arise out

The subsidy game Hypothetical pay-off matrix

[illegible]

		Airbus	
		Produces	Does not produce
Boeing	Produces	Airbus -5 Boeing -5	Airbus 0 Boeing 100
	Does not produce	Airbus 100 Boeing 0	Airbus 0 Boeing 0

[illegible]

		Airbus	
		Produces	Does not produce
Boeing	Produces	Airbus 5 Boeing -5	Airbus 0 Boeing 100
	Does not produce	Airbus 110 Boeing 0	Airbus 0 Boeing 0

Source: *Is Free Trade Passé?* by Paul R Krugman

Krugman proposes, Bhagwati disposes

of a perception of the weaknesses of neo-classical economics and an addition of new insights that go beyond conventional analysis. Instead they typically reflect a failure to understand even the conventional arguments." And he concluded: "I would be surprised if a rigorous attempt to apply the principle of optimal intervention would not actually lead to considerably freer trade than we now have."

The new trade theory needs to be handled with extreme caution for several reasons. Some of these are to do with the politics of trade—notably the danger that the government will be driven by special interests rather than by clever economics. But even judged on its economic merits, the theory has drawbacks.

The most trenchant critic of the application of the new theories to trade policy is Mr Jagdish Bhagwati of Columbia University. Mr Bhagwati is one of the modern pioneers of the classical school, and the author of "Protectionism" (published by the MIT Press), a marvellous brief introduction to the subject. As he tirelessly insists, much of the economic reasoning that has long defended the classical view from infant-industry protection and the optimal tariff is directly relevant to the new (and not-so-new) ideas of the trade revisionists. First, if a cycle of retaliation and counter-retaliation sets in, everybody is worse off. Second, governments lack the information needed to intervene intelligently.

The second point is crucial. To have any chance of making strategic trade policy work, governments would need copious amounts of information. In the Boeing-Airbus case, for example, the interventionist government needs to know how much profit would be earned by going ahead with a new aircraft, both with and without competition from abroad. Small mistakes in such calculations can mean that intervention makes the intervening country worse off, rather than better off.

Economists have shown that in an industry which is a duopoly, the right intervention (if any) might be either an export subsidy or an export tax, depending on the precise mode of competition between the two companies. Moreover, the mere presence of imperfect competition does not make it certain that there is a strategic opportunity to be grasped, even by an all-knowing and omnipotent government; there must also be a lasting supply of excess profits, with no indirect forms of competition to whittle them down. But excess profits are probably less common than governments think. What looks like economic rent to an intervention-minded government may instead be a barely adequate return to justify the risks of the initial investment.

Are there really any industries in which the economies of scale are so great that there is room, as in the imaginary Boeing-Airbus and Japan-America examples, for only one profitable producer? If the global market can sustain, say, four or five producers, then competition will eliminate a lot of the rent that strategic trade policy is designed to capture. A subsidy and/or protection for the domestic producer would then be very likely to reduce national income. On the other hand, while too much competition reduces the rents to be gained, too little may harm domestic consumers: it would be a big mistake to subsidise or protect any old oligopoly for strategic-trade reasons.

Then, once the government is sure that it has weighed all this correctly, it needs to find ways of promoting its chosen industry which do not retard the development of others that might be of equal potential. There are no such ways. To put one industry at an advantage is to put others—initially, at least—at a disadvantage. While the government is conscientiously trying to steer a course through all this, it is beset from all sides by every industry you care to name, pleading its case for special treatment. To conduct a successful policy against such odds would be a miracle.

Brass tacks . . .

But isn't that exactly what America's competitors in the global market have done? This conviction, more than any other, drives the new protectionist sentiment in America. The reality is a lot less clear-cut.

Together with Mr Richard Baldwin of the University of Wisconsin, Mr Krugman has studied the real-life versions, as opposed to the let-us-imagine versions, of the struggle between Boeing and Airbus in the market for medium-range wide-bodied aircraft, and between America and Japan over semiconductors. (Actually, to fit the facts to their economic models, a good many tin-openers have had to be assumed, but let that pass.) These cases, it turns out, are no advertisement for the strategic approach to trade.

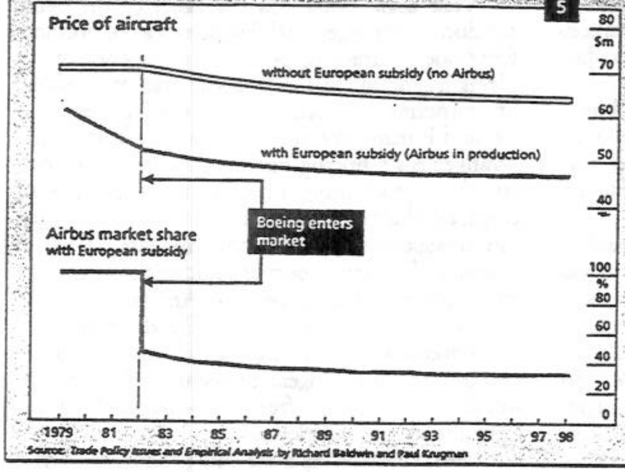
Consider first the Boeing 767 versus the Airbus A300. The authors argue that Airbus received an implicit subsidy from Europe's governments—a view that the governments would challenge. The study* reckons this subsidy took the form of a required rate of return on investment that was lower than that facing Boeing; the authors' model cranks out the implicit value of this subsidy as about $1.5 billion at 1974 prices.

One question is, did this subsidy make Europe better off? The simulation model used in the study says no. Airbus itself lost money on the venture. There was a gain to European airlines and passen-

---

*"Industrial policy and international competition in wide-bodied aircraft." In "Trade Policy Issues and Empirical Analysis", edited by Richard Baldwin, University of Chicago Press and NBER.

Boeing v Airbus **5**



Source: *Trade Policy Issues and Empirical Analysis* by Richard Baldwin and Paul Krugman

Consumer surpluses and profits
$m, discounted to 1974

	Without European subsidy		With European subsidy		Net gain
	Consumer surplus	Profits	Consumer surplus	Profits	
Europe	3,053	0	4,487	(1,471)	(37)
United States	4,892	6,297	7,190	1,020	(2,979)
Rest of world	3,843	0	5,648	0	1,805

() brackets indicate loss

gers because competition in the market kept aircraft prices lower (see chart 5 above). But the amount of the subsidy—in effect, Airbus's discounted losses—roughly matched this gain. Remembering the additional economic costs of the subsidy, which must ultimately be financed through higher implicit or explicit taxes, Europe probably lost. America lost more, though, because Boeing's profits were smaller, and this was not fully offset by the price benefit enjoyed by American aircraft-users. The only clear winners were other countries, whose airlines and passengers benefited from lower prices at no cost to themselves.

On this account, Europe's attempt at strategic trade policy failed. The next question is, what was the appropriate American response? Considering for the moment only the unilateral ("realistic") possibilities, there are three: (a) retaliate cleverly, (b) retaliate stupidly, (c) deter with threats of retaliation not subsequently carried out.

Of these, (c) is clearly the best outcome. If Europe could have been persuaded not to subsidise the Airbus, and the matter had ended there, America would have been a lot better off and Europe (on the Baldwin-Krugman numbers) no worse off. But for a threat to be credible, the threatener must be willing to follow through—or at least be seen to be willing.

That leads back to (a) and (b). If it comes to it, probably the best that clever retaliation can expect to achieve is to make Europe a lot worse off while making America only a bit worse off; this would be the mirror-image of Europe's initial decision to protect. Stupid retaliation could make America a lot worse off, while not harming Europe much—or even helping it. Unless there is a good chance of getting away with (a), therefore, options other than unilateral retaliation would seem to be worth exploring.

... and memory chips

The struggle over semiconductors sheds further light on the matter. In another study* Messrs Baldwin and Krugman examined the competition between American and Japanese firms to produce 16K memory chips—a market that flourished between 1978 and 1984. As in the Boeing-Airbus case, the authors reckon that America faced an unfair trade practice. The evidence for that is rather circumstantial: American firms had a big share of the market both at home and in third markets, but only a small share in Japan (chart 6 on next page). The authors attribute this to implicit infant-industry protection in Japan. On that assumption (which, as in the aircraft case, is open to challenge), they start modelling furiously.

The model says that Japan's protection was equivalent to a tariff of 26%. Because of that protection, Japan had three producers in the industry; under free trade it would have had none. But Japan ended up worse off than it would have been under free trade because its consumers had to pay more for their chips—and there is no offset to that loss from higher profits. (The model plausibly assumes that excess profits are almost zero throughout, because there are always several producers in each market.) To put it another way, Japan succeeded only in allocating resources to a less-than-optimal economic use, at considerable direct expense to its economy as a whole. If this was another case of strategic trade policy, it was evidently another failure.

However, again as before, America lost more from its trade partner's skullduggery than did the partner itself (see table on next page). At the time, understandably, there were many calls for retaliation. Messrs Baldwin and Krugman ask their model what the result would have been if the quarrel had degenerated into a trade war, with each side blocking all 16K-chip imports from the other. The answer is that Japan would have been substantially worse off than under either of the other two scenarios; America would have been worse off than under free trade (obviously), but also a bit worse off than as a supine victim of one-sided protection. Again, retaliation—at least in this form—backfires.

A few years after competition over 16K memory chips had faded, the episode was repeated all over again for 256K memory chips. But with variations. The outcome this time was the Semiconductor Arrangement signed by America and Japan in 1986. America's semiconductor industry had accused Japan of dumping chips in the American market and in third markets, and of restricting its own access to

---

*"Market access and international competition: a simulation study of 16K random access memory". In "Empirical Methods for International Trade", edited by Robert Feenstra, MIT Press.

Subsidies and microchips

	US	Japan
Free trade		
No of firms	7	0
Price of chips $	1.30	1.37
Consumer surplus $m	1,822.5	738.9
With Japanese protection		
No of firms	6	3
Price of chips $	1.47	1.47
Consumer surplus $m	1,651.8	698.4
Net losses in consumer surplus*	9.4	5.5
With US retaliation (managed trade)		
No of firms	7	5
Price of chips $	1.49	2.19
Consumer surplus $m	1636.7	225.6
Net losses in consumer surplus*	10.2	69.5

* As % of consumer surplus under free trade. Figures relate to 1978-84.
Source: "Rethinking International Trade," Paul Krugman

the Japanese market. Japan was asked to stop dumping and to expand its imports of American chips; if it did not, it was left to conclude, it could expect to be bashed with anti-dumping duties.

As agreed, Japan obliged America by raising the price of its chips in the American market. (Between the last quarter of 1986 and the second quarter of 1988, the price of a 256K chip nearly doubled, from $2.10 to $3.80.) But Japan was slow in raising prices in third markets and in "granting better access" to American chip-makers (that is, buying more American chips). So in 1987 President Ronald Reagan announced tariffs on $300m of Japanese imports. Japan raised its chip prices in third markets, and in a few months the sanctions were partially lifted. But America's share of the Japanese chip market still refused to rise.

In judging this example of realistic trade policy, the first question is whether Japan really was dumping its chips. The arrangement looked at the prices of Japan's exports in relation not to prices at home but to "fair market value", defined as average cost plus 8%. The implication is that setting a price of less than this must be an attempt to drive competitors out of the business, with a view to charging monopolistic prices later on. However, the semiconductor industry is one in which sunk costs are high and the variable costs of production are extremely low.

Moreover, learning-by-doing is important: this means that every chip produced makes it, in effect, a bit cheaper to produce the next. So the true cost of producing an extra chip is less than the current variable cost—and that in turn is much less than the average cost. When such an industry encounters a downturn in demand—as did the market for memory chips in 1985 and 1986—the sensible response is to cut prices and maintain output. So even if Messrs Baldwin and Krugman were right about restricted American access to the Japanese market for 16K chips, there is no reason to suppose that the Japanese were dumping (in the predatory sense) 256K chips.

That aside, what was the effect of the Semiconductor Arrangement? The idea was to promote free trade by attacking predatory pricing and by applying a crowbar to the Japanese market. It ended up restricting free trade, helping the Japanese a little, and hurting everybody else a lot. To meet the arrangement's pricing instructions, Japanese firms cut their production, resulting in a shortage of chips, a global rise in prices, and editorials in American newspapers attacking the Japanese for overcharging. That cut in output could stick at a time of rising prices precisely because America had not merely permitted, but required, Japan to create a rent-collecting chip-making cartel. The market-access part of the arrangement—which, if successful, would have promoted free trade—had little effect. Only the stupid part worked.

Meanwhile the Japanese chip-producers carried on "dumping" chips in their own market, thereby giving their electronics industry an edge over the competition in America and Europe. Europe did not take kindly to being a beneficiary of America's tough line with Japan, and complained to the GATT about the arrangement. The GATT found it illegal for the Japanese government to monitor production and export prices in order to restrict exports, and Japan agreed to stop monitoring prices to third countries. That just left increasingly angry American manufacturers to enjoy the fruits of the arrangement.

From strength to strength

Consider one last set of findings on trade policy in an imperfectly competitive world. In a recent article* Mr David Richardson of the University of Wisconsin surveyed studies of trade liberalisation, as opposed to trade intervention. These studies set the main classical objections to trade intervention at nought by, in effect, assuming (however implausibly) that governments can gather all the information they need and use it intelligently, and that there will be no retaliation from other countries. Despite those implicit assumptions, the survey con-

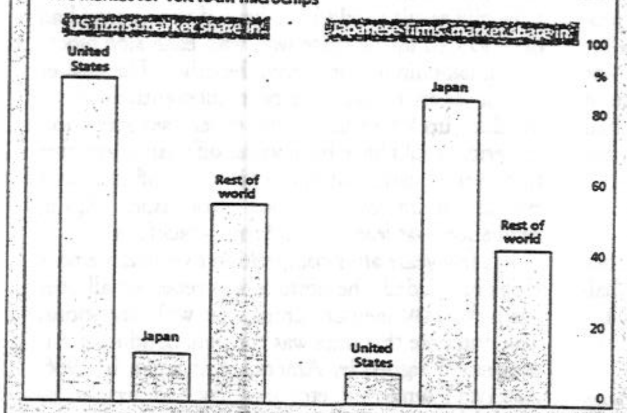
---

*"Empirical research on trade liberalisation with imperfect competition." OECD Economic Studies No 12.

Chip wars

6

The market for 16K RAM microchips



Competitors in the 16K RAM market
1974 $m

Firm	Share of world production, 1977-83
Mostek	15.3
NEC	15.2
TI	12.5
National Semiconductor	10.6
Fujitsu	9.5
Hitachi	6.4
ITT	5.7
AMD	5.4
Motorola	5.4
Toshiba	3.6
Siemens	3.1
Intel	2.4
Fairchild	1.6
Eurotech	1.5
Mitsubishi	1.2

Source: *Rethinking International Trade* by Paul R Krugman cluded that free trade still makes sense. Indeed, imperfect competition seems to strengthen the case for free trade, not weaken it.

The studies have to deal with a balance of pros and cons. To begin with, free trade improves economic welfare in the way the classical theory says (remember the triangles?); this is as true under imperfect competition as under perfect competition. Free trade therefore has a head start over intervention. But the story does not end there. Under imperfect competition, monopolistic firms may be making excess profits. Free trade, by promoting competition, reduces those excess profits. If the firms are making goods mainly for the domestic market, this will probably benefit the national economy (because the excess profits are at the expense of domestic consumers). If the firms are making goods mainly for export, it will probably hurt the national economy (because the excess profits are at the expense of foreigners). The loss in that case might be big enough to outweigh the other gains from free trade. So it all depends.

Of the studies surveyed by Mr Richardson, just four directly ask the question: what is the effect of freer trade on national income, comparing perfect competition with imperfect competition? All four found that the gains from freer trade were bigger under imperfect competition than under perfect competition. One of the most detailed studies looked at the effect of lower trade barriers in Europe after 1992. In ten industries ranging from shoes to machine tools, freer trade yielded a gain in national income in every case—and in nine cases out of ten, the gain was bigger (often much bigger) under the assumption of imperfect competition.

Protectionism has grasped the new trade theory to its bosom. Yet the applied research that this theory has so far prompted says that free trade is not merely best; it is even better than you always thought. ■

Agenda for trade policy

What trade policy should, and should not, aim to do

VERY well, the economics of managed trade is extremely discouraging. Even if a first strike with strategic trade policy meets with no retaliation—the most favourable case—it is unlikely to make the country concerned better off. A similar economic logic applies to retaliation by a country that is already on the receiving end: even if the reply does not provoke counter-retaliation, it is unlikely to advance the aggrieved party's well-being.

But a third fact is politically far more powerful than either of these. If Japan and Europe do subsidise or otherwise promote some of the industries that compete with American firms, America does lose. In the imperfectly competitive world that the new trade theory is concerned with, it can lose in a big way. Understandably, America is likely to insist that something is done about that, even if the result is a self-inflicted wound.

Interventionist instincts run much deeper in Japan and Europe than in the United States. America's traditional belief in the market, and its role over many years as the champion of free trade in the GATT, have served its own interests as much as everybody else's: it is no coincidence that America is the richest country in the world. If trade friction changes America's mind about the role of government in the economy, this will be a disaster for America. But it will be a disaster for everybody else, too, because there will be no powerful government left to fight for the interests of the broad mass of people over narrow groups of producers.

For the moment, America's retreat from the market is almost a celebration. The aggressive new trade policy works, say many in Washington. In 1986, when Portugal and Spain joined the EC, the Community drew its new members into the fond embrace of the Common Agricultural Policy. They were required, among other things, to buy a certain proportion of their grain from other EC countries. The value of American exports affected by the measures was roughly $1 billion a year. President Reagan retaliated with equivalent duties on a variety of EC "yuppie" exports, such as mineral water and expensive cheeses. In due course, the EC bowed to the pressure and negotiated a deal that was satisfactory to the United States.

Super 301 also seems to be working—not least because it has so far been used with intelligence and restraint. As soon as the USTR, Mrs Carla Hills, named Brazil, India and Japan as unfair traders, talks began. Brazil and Japan made trade concessions and soon escaped from the list. Before the list was even announced, Taiwan and South Korea made trade concessions in anticipation of being named. The talks with Japan subsequently blossomed into the Structural Impediments Initiative, under which Japan has made promises about all manner of economic and other policies, aimed at increasing the country's imports from America. Who says realism doesn't pay?

In a recent article intended for popular consumption*, Mr Krugman and two other well-known economists, Mr Rudiger Dornbusch of the Massachusetts Institute of Technology and Mr Yung Chul Park of Korea University made their views clear. Even if there is little to be gained from a strategic struggle over rents (where the stress is on targeting and promoting specific industries), America needs to be aggressive in lowering foreign barriers. The GATT has failed; the bilateral deal is here to stay. Not only that, but trade deals of the old sort, based on agreements to remove barriers, are no use. Replace them, especially in dealing with Japan, with agreements about outcomes. Or, as Mr Dornbusch put it in an article for the *New York Times*: Give Japan a target and say, "Import."

Read my lips

One wonders how America would take to such an instruction from Japan. It is also worth pausing to ask whether Japan is, in fact, an unfair trader. The policy: control protectionist sentiment at home; open foreign markets to imports; help its exporters to compete in those markets; and help the domestic economy adjust to the changes brought by shifting patterns of trade. All four tasks are related, but some of the links matter more than others.

---

*"Meeting World Challenges: US Manufacturing in the 1990s." Eastman Kodak.

GATT-disciplined reciprocity is the way to make the first two tasks simultaneously feasible. Without that discipline, they become incompatible: prising open foreign markets with charges of unfairness, threats of reprisals and "voluntary" import expansions will feed protectionist sentiment and sap the will of domestic firms to compete. That is why it is in every country's own interests to tone down the angry talk, and replace aggressive reciprocity with tolerant reciprocity. Trade policy should be a co-operative effort to promote trade, not a struggle to gain advantage. A market-minded government would have it no other way.

Pursuit of the third goal—helping exporters to compete in foreign markets—also needs to be disciplined. A market-minded government would wish to distinguish carefully between the sort of help that is designed to capture rents (strategic trade policy) and the sort that is designed to improve the allocation of resources (by taking account of externalities). Such a government would be more than willing to deny itself the former, but would expect other governments to do the same; it might want to reserve the right to adopt the second sort of policy, but if so it would be willing for other governments to do likewise. Some policies would fall clearly into either the prohibited category (export subsidies, for example) or the allowed category (government spending on basic R&D, for instance). Others would be harder to place. That is precisely why, as before, an impartial judge must be part of the system.

Similar considerations apply to the fourth goal—helping parts of the economy that come under trade-related stress. A market-minded government might want to grant temporary, transparent protection (with a pre-announced rate of withdrawal) to industries that lose out in the market as the pattern of comparative advantage shifts. While it lasted, this protection, obviously, would distort trade a bit. But the government would ask for no exemption from the system's disciplines that it was not willing to grant to others in the same circumstances. Again, what counted as an exemption from the rules, once agreed, would need to be decided by an impartial outsider.

In this way, a multilateral, rule-based system liberates market forces, leaving them to determine patterns of trade. This is what the GATT is for. If market-minded governments only choose to use it, the GATT can help them to achieve all four of these legitimate trade-policy goals. ■

Mrs Hills and her Super-301 six-gun

Meanwhile in Geneva

Can the trade negotiators meet the challenge?

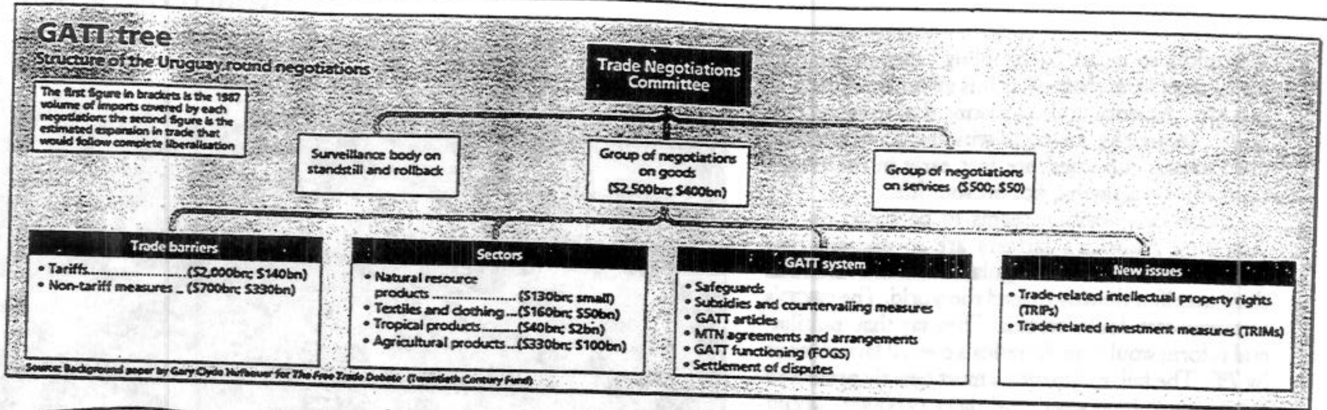
EVEN under favourable circumstances the Uruguay round would be difficult enough, because it is immensely complicated. But the circumstances are anything but favourable. As in previous rounds, America is the driving force: it has come to Geneva with a big list of new rules that it wants to see written into the GATT. In particular, it has repeatedly threatened to walk away from the talks if the EC refuses to make real headway on liberalising the CAP. The other GATT members all have their own priorities, but their biggest common interest is in wanting America to leave them alone—to stop bashing them with countervailing duties, anti-dumping cases and both varieties of Section 301.

At first sight, the room for compromise seems small indeed. Time is short (though the round could be extended for at least a few months beyond its December deadline without too much difficulty) and precious little has so far been agreed. The outlook, then, is hardly promising.

The negotiations are going on in 16 different groups. The box on the next page lists them, together with the broad negotiating goals which have already been agreed. Courtesy of Mr Gary Hufbauer of Georgetown University and the Institute for International Economics, the box also gives an estimate of (a) the volume of trade falling into each category and (b) the expansion of trade that would happen if each group agreed to liberalise the trade in its jurisdiction completely. Agricultural trade is the most important single category, for two reasons. First, complete liberalisation there would expand trade by some $100 billion a year—substantially more than any other area might yield. Second, America has decided that it will be the make-or-break issue in the round.

Down on the farm

America's insistence on radical reform in agriculture is fully justified. But note the irony, too. Farming offers a glimpse of just how dreadful managed trade between supposedly free-market economies can be. Agriculture was exempted from GATT disciplines (at America's request, by the way) in 1955. The outcome is a warning to those who would do the same with other sorts of trade. Many commen-



GATT tree

Structure of the Uruguay round negotiations

The first figure in brackets is the 1987 volume of imports covered by each negotiation; the second figure is the estimated expansion in trade that would follow complete liberalisation

Trade Negotiations Committee
Surveillance body on standstill and rollback
Group of negotiations on goods ($2,500bn; $400bn)
Group of negotiations on services ($500; $50)
Trade barriers
• Tariffs........................($2,000bn; $140bn)
• Non-tariff measures _ ($700bn; $330bn)
Sectors
• Natural resource products ....($130bn; small)
• Textiles and clothing ...($160bn; $50bn)
• Tropical products.........($40bn; $2bn)
• Agricultural products...($330bn; $100bn)
GATT system
• Safeguards
• Subsidies and countervailing measures
• GATT articles
• MTN agreements and arrangements
• GATT functioning (FOGS)
• Settlement of disputes
New issues
• Trade-related intellectual property rights (TRIPs)
• Trade-related investment measures (TRIMs)

Source: Background paper by Gary Clyde Hufbauer for *The Free Trade Debate* (Twentieth Century Fund)

tators say that agriculture is the GATT's biggest failure—and so it is. But in another way it is also the best advertisement the GATT could wish for.

Economists at the OECD (the Paris-based club of rich-country governments) have recently studied the costs of agricultural protection in some detail. Although farming accounts for only a small part of the output and jobs of most industrial economies (almost everywhere, less than 5% of both), it is so badly run that the costs of protection are a substantial burden on the whole economy. The OECD calculates that incomes in these economies would rise by almost 1% on average if all farm protection was removed (chart 7). In other words, agricultural support costs the OECD countries about $72 billion a year (at 1988 prices). That is what economists call a deadweight loss: the gross costs to consumers (in taxes and higher prices) may be three times as great.

To reap the gains of reform, less than 1% of the workforce would need to be displaced. The cost per farm job "saved" is roughly $20,000 a year in lost efficiency alone; this excludes the direct cost of income transfers to farmers. And this is not the cost of increasing employment by one job, it is the cost of having one more job in farming instead of one more job in some other industry.

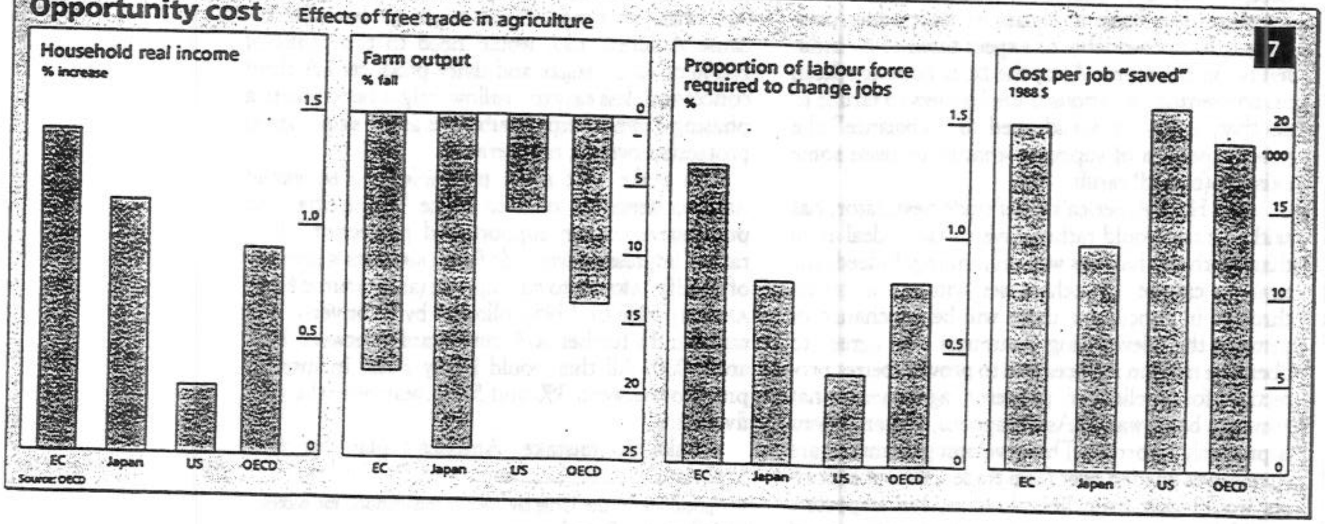
Worldwide, the costs of farm protection are even greater. The OECD's estimates ignore farm protection in some of the smaller OECD countries. Also, they take no account of the fact that farm liberalisation would raise the incomes of many developing countries, by enabling them to export more farm produce in a free global market. One estimate puts this potential increase in third-world agricultural exports at nearly $30 billion a year.

It is worth noting that the individual countries would all gain from liberalising their farm policies unilaterally. Indeed, most countries gain more from reform if they go it alone than if everybody joins in. The main reason is the terms-of-trade effect (this was explained earlier, in connection with the optimal tariff). Multilateral farm liberalisation would make the industrial countries' exports cheaper relative to their imports. This offsets some of the income gain that they would otherwise enjoy. For instance, chart 7 shows that incomes in the EC would rise by 1.4% if all countries liberalised their farm trade. If the EC dismantled the CAP and nobody reciprocated, its income gain would be 1.6%.

From the point of view of any individual country, the case for multilateral reform is, as always, a largely political one. Multilateral reform allows each government to say that farmers abroad are suffering as well. This is not entirely a matter of presentation. By spreading the burden of adjustment around, multilateral liberalisation slightly reduces the fall in farm output that each liberalising country has to face.

A way to a deal

Multilateral farm reforms are clearly in (almost) everybody's interests. Is there any chance that the GATT talks will deliver them? America wants a clear commitment to free trade in agriculture. This would mean converting quotas and other direct controls into tariffs, and then reducing those tariffs quickly.



Opportunity cost

Effects of free trade in agriculture

7

Source: OECD

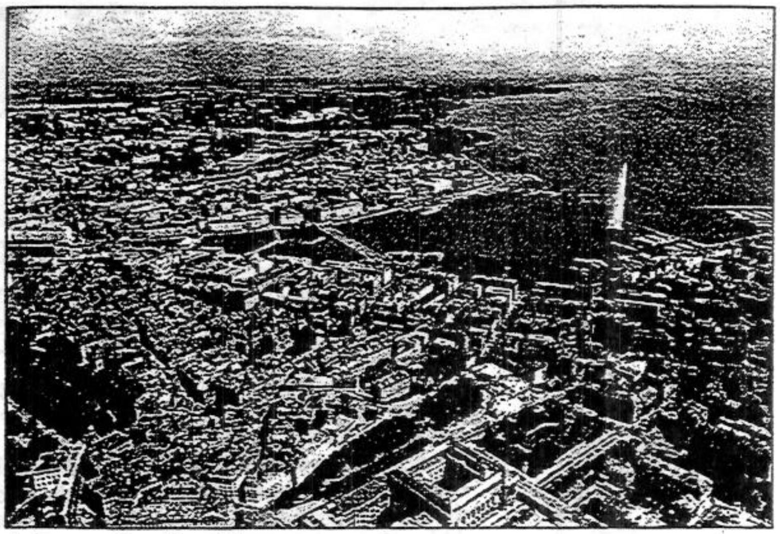
It would also mean "decoupling" income support for farmers from their decisions about how much to produce. America is also insisting that some specific targets (especially the rapid winding down of European export subsidies) are met early in the transition. The EC is saying, not on your life.

America's position seems to be based on the view that America's farmers are so efficient that they will thrive in an unregulated market, and fulfil their manifest destiny to feed the world. The OECD's numbers cast doubt on this. They say that multilateral reform would cut America's overall farm output by 7%. The fall in America's most heavily protected sectors (including sugar and dairy products) would be much steeper. There must be some doubt that Congress, if the time ever came, would agree to put America's negotiating position into practice. (If recent versions of the farm bill are any guide, Congress has its own ideas about agricultural protection.) Judged on their merits, however, America's proposals on agriculture make excellent sense. And America has valuable support in the talks from the Cairns group—an alliance of rich and poor agricultural exporters that would benefit from freer farm trade.

The EC's position is more complicated. The Community sees America's proposals, which stress border measures (such as import levies and export subsidies) over domestic support for farming, as an unbalanced attack on the particular sorts of protection the CAP provides. In this, America has economics on its side: the CAP's border measures do more economic harm than America's expensive system of payments to farmers. In particular, American consumers buy more of their food at close-to-market prices than do Europe's consumers. Politically, America has mapped out a route to free trade in agriculture that initially puts a far heavier burden of adjustment on Europe than on itself.

Europe's counter-offer, broadly, is to propose a route to less protection (not free trade) that is based on an aggregate measure of support (AMS). The idea is to capture all the different forms of protection in a single measure and agree upon targets for how this single measure should be reduced, thereby leaving it to governments to decide which measures to cut and by how much. This might seem fair enough, except that the choice of which measures to liberalise does have implications for trading partners; it is unreasonable to expect them to be indifferent on the point. Also, the EC is none too keen on converting its various trade barriers to tariffs; to do that, it says, it would need to "rebalance" the present pattern of support—that is, increase some existing (bound) tariffs.

Mrs Hills, America's chief trade negotiator, has said that she would rather have no GATT deal at all than let the EC have its way on farming. Indeed, she says, it cannot be otherwise: without a breakthrough in agriculture there will be no chance of getting the developing countries to agree to liberalise trade in services and to provide better protection for intellectual property, agreements that America badly wants. As it happens, this argument is probably incorrect. The developing countries are themselves divided over farm trade: efficient exporters would gain from liberalisation, but importers would lose (because an end to oversupply would raise the prices of many farm goods). It is America, not the third world, that is stamping its foot over farm trade. In the end, "success" will mean finding an agreement that America and the EC can live with.



A funny place for a Uruguay round

Of many possible compromises, Mr James Rollo of Britain's Royal Institute of International Affairs* has suggested one that might fly. Its elements would be as follows. The EC would drop the idea of casting agreements in terms of an AMS. Instead it would make commitments on specific policies (including export subsidies). An AMS, such as the OECD's producer-subsidy equivalent, could still play a useful part in assuring a roughly equal pace of liberalisation on all sides, and in guiding the parties on how much was still needed to be done in later rounds of reform. The EC's other concession would be to accept a firm timetable for the conversion of non-tariff measures to tariffs. This timetable might start once the agreed cuts in support had been made. (Converting to tariffs before the first round of cuts would pose difficulties in any case, because the tariffs would have to be very high—more than 200% in some cases.)

In return, America would save for next time its insistence on an explicit target of zero support and protection. It would also have to accept a delay in the switch to tariffs. (That might be bearable because America, too, would need to put tariffs of 200% or so on sugar and dairy products.) A third concession, less easy to swallow, might be to accept a phasing-down of export subsidies at the same rate as protection overall, not faster.

To make that more palatable, the EC would have to define its offer to make "substantial and progressive cuts in support and protection" in a rather impressive way. Mr Rollo suggests a promise of specific reforms to cut support (as measured by an AMS) by 50% by 1995, followed by a conversion to tariffs and a further 50% cut in tariffs between 1995 and 2000. All this would imply a fall in internal prices of between 3% and 5% a year over the next five years.

Make no mistake, America's plan for farm trade, whether offered sincerely or not, is the right one. It is overwhelmingly in the interests of Europeans that the the EC should surrender to American pressure. Unfortunately, that is unlikely to happen. Mr Rollo's compromise is less attractive than the American plan. Nobody would be entirely satisfied with such a deal. However, it might be feasible; it represents a huge improvement over the current state of affairs; and it leaves the rest of the round to play for.

---

*"Agriculture in the Uruguay round: foundations for success." RIIA Discussion Paper 26.

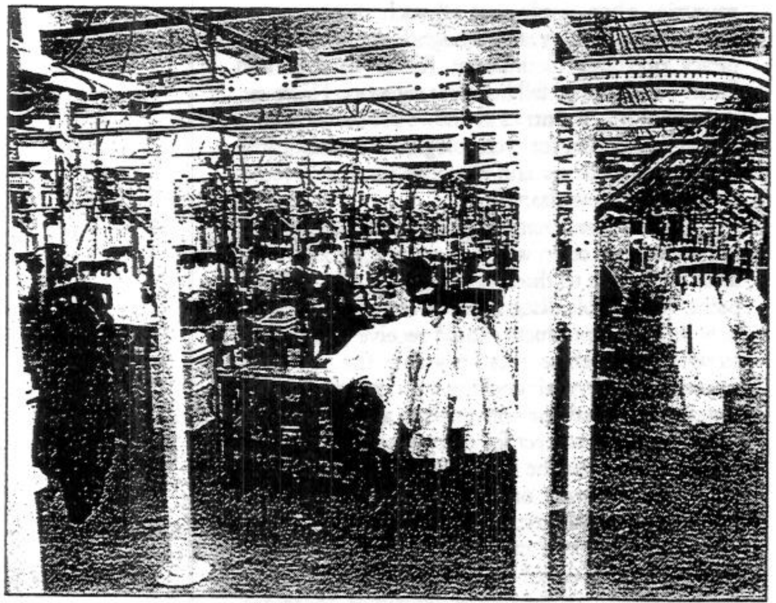
Something old, something new

In textiles and the "new issues", rich and poor countries are in opposing camps

IF AGRICULTURE is the best example of managed trade between supposedly market-based economies, textiles and clothing run it a close second. World trade in textiles is conducted through what Soviet economists would doubtless call a "regulated market"—a web of bilateral quotas that tell each of the exporting countries how much they are allowed to sell to each of the importing countries. The system is called the Multi-Fibre Arrangement (MFA). The exporters are mainly in the third world, and include some of the poorest countries in the world, such as Bangladesh; the big importers are the rich industrial countries.

The MFA was set up in 1974, an all-encompassing version of several earlier deals. It was intended to provide temporary protection to developed-country producers, whose business was threatened by a surge of cheap imports. The GATT went along. As is the way, the arrangement has been extended three times, and is more restrictive now than ever before. A study by Mr William Cline of the Institute for International Economics estimates that free trade in textiles and clothing would increase the industrial countries' imports by $50 billion a year (though the benefit to developing-country exporters would be somewhat offset by a loss of economic rent). Needless to say, consumers would gain through lower prices. Even rich-country producers are a bit less keen on the MFA than they used to be: automation has reduced their costs, and they have moved their business up-market.

This may be why some GATT-watchers believe that real reform may be possible in the Uruguay round. Hitherto, it had been taken for granted that the MFA would survive; the only question was the speed at which it would become more restrictive. At last there seems to be a consensus that liberalisation is overdue.


Finding a niche up-market

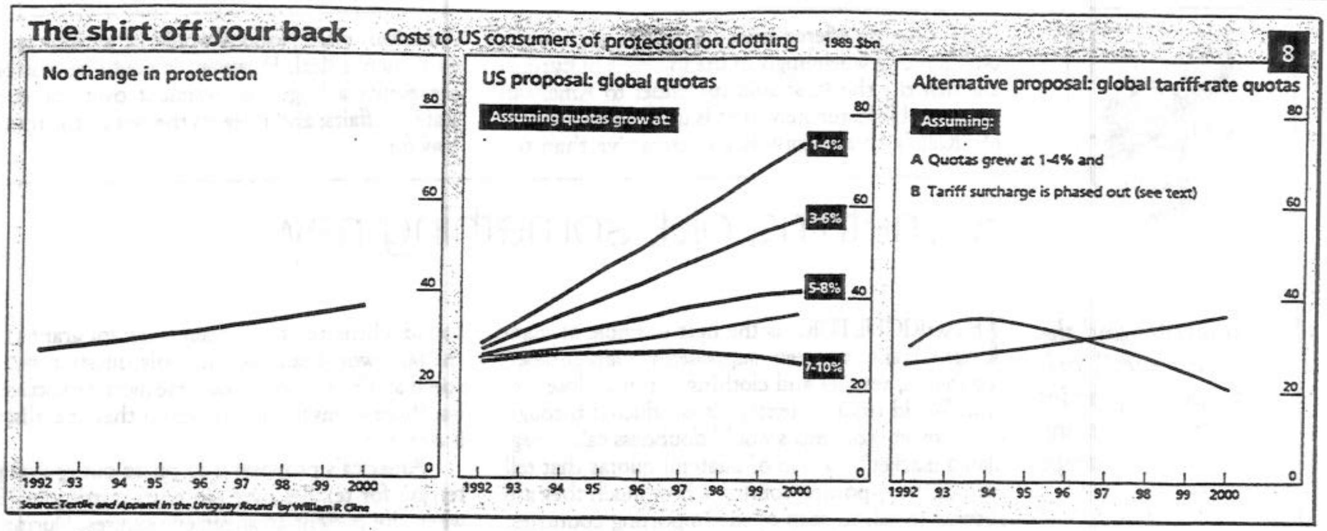
America's proposal is to phase out special protection for textiles over ten years starting in 1992, when the present arrangement expires. During the transition, America wants to replace the MFA bilateral quotas with a global quota for each product. To begin with, this global quota would consist of the existing bilateral quotas plus a non-selective "global basket". Each year the basket would grow at a rate to be agreed; America suggests some minimum growth rate for the first three years, the minimum plus one percentage point for the next three, the minimum plus two for the next two, and the minimum plus three for the last two. At the same time, and in addition, the existing quotas would gradually be reallocated to the basket. So after ten years there would be a single, enlarged quota for each product, which importers would be free to allocate to the producers of their choosing. At that point, the global quota itself would be abolished.

All in the fine print

The plan has its attractions—both for free-traders and for protectionists. For a given level of overall protection, a global quota is less harmful than a set of bilateral quotas, because it allows importers to buy from the cheapest source. With time, this would lead textile production to be concentrated in the most efficient countries. On the other hand, America's scheme has the potential to be more protectionist than the present one. In effect, it extends quota restrictions to the (mainly industrial) countries that are not at present covered; that is why the EC is against the idea.

In the end, as economic simulations by Mr Cline make clear, everything would depend on how quickly the global quotas were expanded. First, Mr Cline works out the costs to American consumers if the present system of protection for clothing continues indefinitely. On this assumption, his model reckons that America's imports could grow at around 7% a year. Even though the level of protection is assumed to be unchanged, the absolute costs to consumers rise along with the expansion of demand and the size of the market (more goods are changing hands at distorted prices). In this baseline case, therefore, the costs (in 1989 prices) rise from $26 billion in 1992 to $37 billion in 2000 (see chart 8, left-hand panel, on next page).

Mr Cline next asks his computer what the costs of protection would be under a global-quota scheme, assuming a variety of growth rates for the quota. If the growth rate was 1% in the first three



The shirt off your back Costs to US consumers of protection on clothing 1989 $bn

Source: *Textile and Apparel in the Uruguay Round* by William R Cline

years of the transition, rising (as under America's sliding-scale plan) to 4% in the last two, the consumer costs would more than double by 2000 (chart 8, centre panel). The global quotas would need to expand by 7% a year rising to 10% a year merely to cap the costs of protection. And that, rather than providing a smooth transition to free trade in 2000, would still leave an adjustment shock as great as now. Quota growth of 8% rising to 11% would be needed to cushion the eventual switch to free trade. Anything less would cast doubt on the promise to scrap protection.

Such rapid expansion of the quotas will be politically difficult, to put it mildly. While America's negotiators in Geneva stressed their liberal intentions, Congress was debating a new textiles bill that proposed, in effect, a global quota with 1% growth throughout. The clash between politics and economics is starker here even than in farming. If such a measure ever became law, it would cost consumers another $48 billion a year by 2000, "saving" roughly 250,000 jobs (relative to the baseline) at an annual cost of more than $190,000 apiece.

The same political difficulty applies, only more so, to the developing countries' proposal to liberalise by quickly expanding the existing MFA quotas to the point where they no longer impinge on trade. This would imply annual quota growth rates reaching 20% and more within five years.

Mr Cline favours another approach, which has also been offered as a possibility by the American GATT negotiators. This combines economic sense with—he hopes—greater political acceptability. The idea is to accept a form of global quota and to let it grow comparatively slowly. However, instead of simply banning imports above this quota, the government would make importers pay a higher-than-usual tariff on them. Imports within the quota would come in at the standard tariff (22.5% in America); imports above the quota would face an initially heavy surcharge of, say, 40%.

To remove the MFA smoothly over ten years, the quota could be allowed to grow at 1% a year rising to 4%, and at the same time the surcharge would be gradually phased down to zero. This would actually give importers a bit more protection over the next few years than they could expect in the baseline (chart 8, right-hand panel). Later, as the surcharge faded away, the market would open up. The negotiators in Geneva would do well to study Mr Cline's calculations.

Modern times

Developing countries would be big gainers from any substantial reform of the MFA. More perhaps than in farming, a breakthrough in textiles and clothing would be seen by them as a concession that warranted a spot of reciprocity. The industrial countries, led of course by America, have left the developing countries in no doubt about the measures they would expect in return. The Uruguay round includes talks on a range of trade issues that have hitherto fallen outside the scope of the GATT. America regards three of these "new issues" as especially valuable. Unless they win some sort of agreement here, America's negotiators will regard the round as a disappointment.

- **Intellectual property.** Many producers in Europe and America complain that their patents, trademarks and other intellectual-property rights are infringed in foreign markets, especially in the developing countries. This debate suffers from an excess of self-righteousness. Companies in rich countries often imply that all such disputes over intellectual property are a straightforward matter of piracy or theft. Counterfeit Rolex watches and Gucci handbags do fall into that category. However, the developing countries are more concerned about patent protection for technologically sophisticated goods such as drugs and chemical fertilisers. Here, the issues are less clear-cut.

In many cases, matching rich-country standards of patent protection would make such goods a lot more expensive to third-world buyers. More to the point, respectable disagreements are possible about how long new products should receive patent protection. This debate goes on within the industrial countries; yet, in their dealings with the third world, companies regard the conventions agreed at home as self-evidently correct for everywhere else. It is not at all obvious that the developing countries are obliged, either morally or for the sake of sound economics, to meet the rich countries' demands.

Concerns about piracy and patent protection have fuelled the use of procedural protection, especially in America. Despite the difficulties, therefore, the GATT's negotiating group on trade-related intellectual property rights (TRIPS) needs to make some progress on the matter. The parties still disagree even about the broad approach that ought to be taken—whether it should be a single comprehensive deal, to be written into the GATT itself and subject to the normal dispute-settlement procedures, or a loose bundle of smaller agreements, with tough rules on trade in counterfeit and pirated goods, but leaving governments greater discretion on patent protection and other matters.

- **Foreign investment.** The GATT's group on trade-related investment measures (TRIMS) is looking at a list of 12 policies that interfere with companies wishing to invest abroad. The most restrictive of these measures are rules about local content (the company's foreign plant must buy some minimum proportion of its materials locally), export performance (a given share of output must be for export) and foreign exchange (for instance, if the company's operation will need imported inputs, they must be paid for with earnings from exports). According to one estimate, roughly half of America's recent direct investments in developing countries have had local-content or similar strings attached.

Progress in the TRIMS talks has been slow, mainly because America, which initially pressed for investment measures to be included in the Uruguay round, seems to have lost some of its passion on the subject. Its firms appear to be not too bothered by such rules, because they are generally offered investment sweeteners such as favourable tax treatment in return. Equally, America's government has itself dabbled with local-content rules for inward investors. If the recent controversy about foreign ownership of American assets continues (remember the fuss when a Japanese investor bought the Rockefeller Centre?), America might want to deploy more TRIMS of its own.

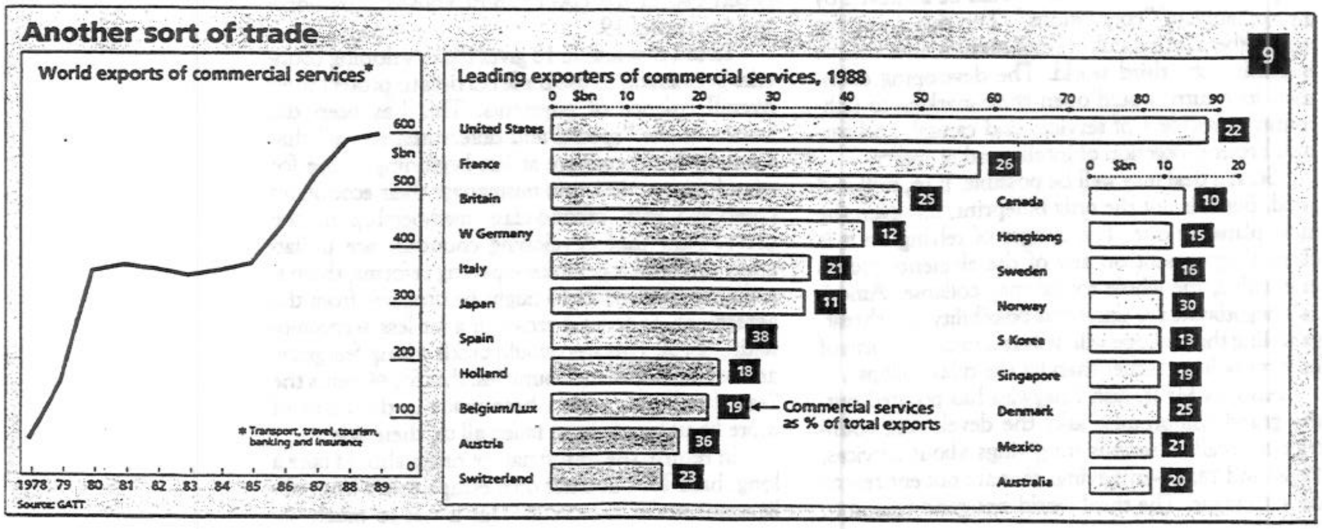
Officially, at any rate, America and the other industrial countries do want to establish some disciplines on the use of investment measures. But many developing countries are against the idea on principle: to tie their hands in this area would be an infringement of sovereignty over essentially domestic policy, they say. Any breakthrough in this part of the round will come as quite a surprise.

- **Services.** This is much the most important of the "new issues". Trade has traditionally been shorthand for trade in goods. But industries such as banking, insurance, telecommunications, transport, tourism, construction, consultancy and so on account for well over half the national income of most rich economies. Nearly 70% of America's GDP and three-quarters of its employment are in services. Lately, international trade in services has been growing, too. Services account for more than 20% of America's exports, and for about the same proportion of exports worldwide.

Many developing countries are themselves big suppliers of services to the world market (chart 9 below). However, most were initially against including services in the Uruguay round at all. Countries such as Brazil and India feared—not without reason—that rich-country banking and insurance firms would steal business from their own less efficient domestic providers, whereas the rich countries would bar the third world from exporting the services in which it has its greatest comparative advantage. How willing will America be, for example, to let teams of low-paid foreign construction workers compete with its own hard hats?

Despite such reservations, services were included in the round. Better than that, the outlines of a worthwhile deal are emerging. It seems the negotiators will agree to a deal that will apply both to services that are supplied across borders and to those that are supplied by a foreign company operating within the domestic market (which would usually imply cross-border movements of capital and labour). Countries will be able to rule out certain services in this first deal; later rounds of talks would aim to remove such barriers. The agreement will facilitate that by embodying the main GATT principles of reciprocity, non-discrimination and transparency.

The main worry is that America, having pressed so hard to get services into the round, will now want to exclude some big industries from the deal. Its shipping and aviation lobbies have a lot to lose. Shipping is protected by the Jones act (which, nothing if not thorough, reserves coastal traffic for American-built, American-owned, American-registered and American-crewed vessels), and aviation is quite happy with the international system of bilateral landing rights that is already in place. America's


Another sort of trade

World exports of commercial services*

* Transport, travel, tourism, banking and insurance

Source: GATT

Leading exporters of commercial services, 1988 banks are keen to make the developing countries' nightmares come true by thriving in new foreign markets, but America's Treasury has doubts. It fears that prudential supervision might be weakened in the process. If these American industries are excluded, other governments will doubtless insist on further exclusions to suit their own lobbies.

Even if the recent momentum can be maintained, many questions both of broad principle and of mind-numbing detail will remain to be answered. How many services will countries be required to offer as their "entry fee" in this first round of liberalisation? How much of a discount on this fee will the developing countries be offered? What will be the mechanics of dispute settlement? Will the framework agreement incorporate a negative list of national measures to be excluded from liberalisation, or a positive list of measures to be liberalised in due course?

Special treatment for the developing countries is likely to be the main sticking point. America's negotiators are arguing that freer trade in services will be so beneficial to the developing countries that no extra concessions or special rules should be offered, just more time to dismantle the present barriers. That view, a striking reversal of America's prejudices about many of its own trade restrictions, is quite correct. Many developing countries seem to agree, and not just those that rank among the top services exporters worldwide, as some do. Developing countries with little scope for exporting services put their goods exporters at a big disadvantage if they force them to pay well over the odds for locally provided financial services. Now that many developing countries are starting to look more kindly on trade liberalisation than before, the point is well understood.

The consensus in Geneva is that with or without the third world, there will be a GATS (General Agreement on Trade in Services) before long. With any luck, though, the developing countries will not only enter this tiresome new acronym, but will do so as (fairly) equal partners, with (fairly) equal rights and obligations.

Once and future GATT

Beyond the Uruguay round, will the world need a new trade watchdog?

DISCUSSION of the terms on which the developing countries might join the new agreement on services is all too familiar. That the third world has not, as yet, been a full and equal partner in the GATT goes to the heart of two critical questions about the future of the global trading system. The first is whether, over the next few months, the negotiators in Geneva can construct a deal to refute the advocates of managed trade. The other is for the longer term. Suppose the round is a genuine success, and further suppose that an open trading system is both feasible and desirable. Is the GATT the right way to take this project forward?

Since the Punta del Este declaration of 1986, which made it clear that the Uruguay round would ambitiously try to cover a broader range of trade issues than any previous round, GATT-watchers have talked of the grand package that would span the different negotiating groups. This deal was to employ the productive fallacy that the GATT has long relied upon: freer trade would be achieved by an exchange of "concessions". The rich countries would liberalise agriculture and textiles, thus mainly benefiting the third world. The developing countries, in return, would open their markets to rich-country providers of services and capital, and furnish better protection of intellectual property.

Such a deal may well be possible. If so, well and good. But it is not the only blueprint, nor even the most plausible one. The danger of relying on it is clear: if agreement on any of the elements proves impossible, the whole round may collapse. America's negotiators are using this possibility as a threat, gambling that Europe will deliver a radical reform of its farm policies rather than let the talks collapse.

Also, as Mr Jagdish Bhagwati has pointed out, the grand compromise asks the developing countries to swallow serious misgivings about services, TRIMs and TRIPs—misgivings that are not entirely to do with trade. The third world has good reason to regard these issues as outside the scope of the traditional give-and-take on trade policy. On top of all this, delivering the grand compromise will be immensely complicated, precisely because it cuts across the negotiating structure of the talks.

Mr Bhagwati and others therefore suggest an alternative. Instead of aiming for a deal that depends on the bundling together of new and old issues, the negotiators should adopt a two-track approach.

Articles of faith

Track one would be a deal on trade in goods. It is here, in the GATT's traditional domain, that the threat to liberal trade is greatest. Reform in agriculture and textiles, along the lines suggested earlier, would be immensely beneficial in itself and would do much to restore the GATT's credibility. But failure to achieve real reform remains a possibility, especially in farming. In that event, a worthwhile deal could still be done. It would involve a trade-off between two of the GATT's most abused provisions: articles 18 and 19.

Recall that article 18 gives the developing countries permission to use trade barriers to protect their overall balance of payments. This has been the source of the "special and differential status" that many governments are at last beginning to see for what it is: permission to mismanage their economies combined with second-class membership of the GATT. Now that developing countries are unilaterally undertaking market-opening reforms, there is a chance that article 18 might be dropped from the agreement, or at least recast in a far less accommodating form. This step would tackle a long-felt grievance of the industrial countries: that for 40 years the "sacrifices" which have been made in the name of more liberal trade have fallen all on their side.

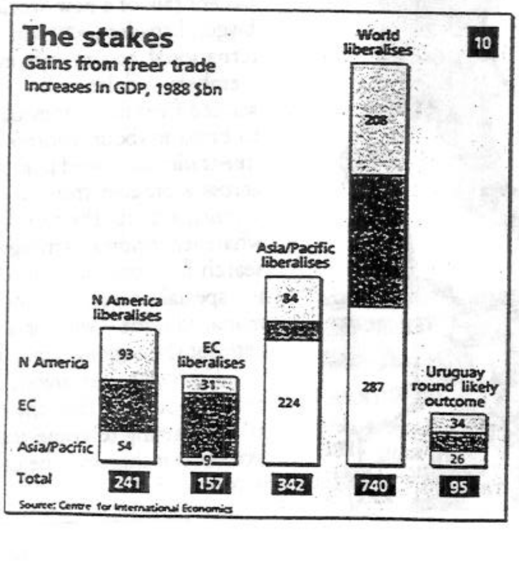
In return, the industrial countries should take a long, hard look at their own escape route from obligations under the GATT. This is not so much be-

For example

cause of article 19, the safeguards clause, which was designed to discipline their use of trade protection, but rather because of their eagerness to find ways round it. The VERs and other non-tariff barriers that have proliferated in recent years are all, in the spirit if not the letter, violations of article 19. These new methods of protection are discriminatory, opaque and far from temporary. More than any other recent development, their growing use undermines the liberal trading order established in the postwar years.

In return for the abolition of article 18, the industrial countries should therefore promise to submit to the discipline of article 19. Instead of VERs and the rest, they should stick to temporary, non-discriminatory tariffs. The result would be less protection, especially of the most damaging varieties. All sides would benefit, as with any move towards more open markets. But this particular deal would have the added advantage of announcing a joint determination to liberalise trade co-operatively rather than confrontationally, and to use the GATT to do so.

The second track would be to look for deals, where they can be made, on the new issues. Agreement to bring services into a GATT-like framework that would encompass the developing countries looks achievable. Intellectual property and foreign investment will be trickier. A possible way forward in both cases might be to adopt what GATT-watchers call the "codes" approach. This would bring TRIPS and TRIMS into the GATT, but would extend the agreement's rights and obligations only to the members that signed up for them. It would bring an element of discrimination into the GATT: some countries would benefit from discipline on investment measures, say, while others would not. But this is a much less corrosive form of discrimination than the article 18 variety, which offers some members rights without matching obligations. In due course, the non-signers might wish to join in, and would be free to do so.



Source: Centre for International Economics

The free-traders' lament

In the beginning, the GATT was a smallish group of fairly like-minded countries. Accordingly, the idea of a co-operative march to more liberal trade was plausible. As the membership grew, co-operation became more difficult. A north-south division emerged, sanctified not just by article 18 but in other ways, too. It began to seem that the GATT was divided between countries that wanted freer trade and countries that did not. The time had come, it seemed, for the free traders to start a new club. The codes approach, on this view, is a holding measure: a series of little clubs within the old club. If it is applied to too many issues, the GATT will simply break down. Therefore, create a super-GATT, comprising only those countries that really believe in free trade.

This is an attack on the GATT, as it were, from the other side—not from the protectionists, but from the free marketeers. It gains strength when you consider the benefits to the world of a not-too-disappointing Uruguay round, and compare these with the benefits of a stronger dose of trade reform.

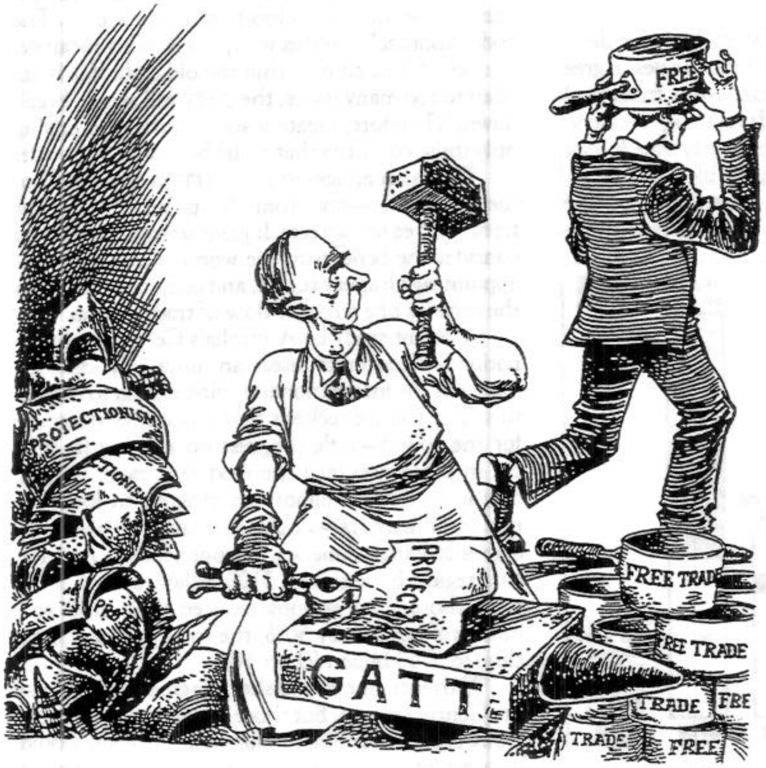
A recent study by Australia's Centre for International Economics used an unusually detailed model of the international trading system to do just this. The Centre reckons that a plausible outcome for the round—little progress on farming (a 10% cut in protection over the next ten years), modest cuts in tariffs, but almost no progress in lowering non-tariff barriers—would increase the joint incomes of North America, Europe and the Asia-Pacific region by less than $100 billion a year. If instead those three regions reduced their tariff and non-tariff barriers by 50%, the total gain would be close to $750 billion (chart 10).

With luck, the round should achieve more than the Centre expects. But those numbers are striking. If any region liberalised unilaterally, its gains would be substantially bigger than the number pencilled in

Jaw, jaw is better than war, war

for the GATT's attempt at multilateral liberalisation. This suggests that the regional approach to trade which is already under way—Project 1992 in Europe, America's free-trade deal with Canada, to be followed perhaps by another with Mexico—may be the way forward. Let the like-minded pursue free trade where they may, and leave the GATT to fade quietly away.

This would be a big mistake. The view that the developing countries will frustrate any bold, multilateral attempt to liberalise world trade is out of date. The third world's ideas on trade are changing. All of a sudden, despite the new regional trade pacts, the doubters and foot-draggers are in the rich, industrial north.



Overall, America has substantially raised its protectionist barriers—and barriers of the most damaging kind—during the past decade. In Europe, Project 1992 seems certain to promote freer trade within the EC, but a willingness to promote freer trade with outsiders is still in question. As one of the European Commission's top officials famously (and fatuously) put it, "The EC sees no reason why the benefits of its internal liberalisations should be extended unilaterally to others." And the Community has failed one early test abysmally: it looks likely to negotiate a set of "transitional" VERs for Japanese cars, copying one of the worst errors in the recent history of American trade policy.

The fact is that no institutional arrangement can make up for lack of conviction where it counts. Recent talk of a new World Trade Organisation—a bigger, broader GATT with teeth—is fine, but no international accord can ever be stronger than its members will let it be. If governments genuinely wanted freer trade, they could use the existing GATT to bring it about. Moreover, in that case, regional free-trade deals need pose no threat to liberalisation across a broader front. If, on the other hand, governments doubt the case for free trade, they will use whatever trading arrangements are at hand to search fruitlessly for strategic advantage, or to satisfy special interests. At the end of this road lies mismanaged trade—with or without the GATT, with or without the regional trade blocks.

The GATT has always been a good tool—but only a tool. Now that the developing countries are at last starting to come round to free trade, it can work better than ever before. This is no time to cast it aside.

23

WHARTON REPROGRAPHICS

WORLD TRADE

Time for another round

The world's current economic troubles strengthen the case for a new round of world trade talks, says Patrick Lane

AT A birthday party in Geneva, revellers are celebrating the 50th anniversary of the General Agreement on Tariffs and Trade (GATT), and a local television crew is making a documentary about the World Trade Organisation (WTO), the GATT's successor. The film might have featured grey-suited trade ministers congratulating each other on 50 years of trade liberalisation, but instead the camera captures a colourful crowd of demonstrators on the far side of rolls of barbed wire. They denounce the organisation and all its works. Free trade, they claim, despoils the environment and enslaves dispossessed peoples. "God is dead," reads one banner; "The WTO has replaced Him." Most of them protest peacefully, but a few start throwing stones and bottles, then overturn cars and set them ablaze.

It would be easy to dismiss the protesters as cranks. After all, the WTO has plenty to celebrate. Global trade has grown 16-fold since 1950 (see chart 1, next page), far outstripping the growth in GDP. It is freer than at any time since before the first world war, and possibly freer than ever. The WTO is an increasingly popular club: its membership now stands at 132 (see chart 2, next page), with another 30-odd, including China and Russia, in the queue to join.

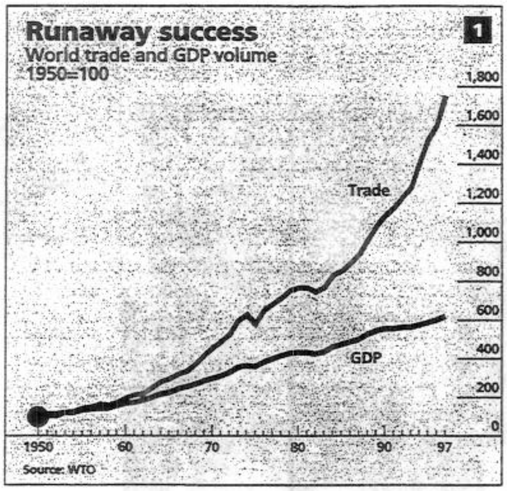
But the free-traders are not having it all their way. President Clinton has failed to win "fast-track" negotiating authority from Congress for new trade agreements, and is now distracted by other matters. Much more importantly, the global economic outlook has deteriorated from the merely worrying to the critical. Already the economies of Japan and much of Asia are shrinking. America, having enjoyed a prodigious boom for the past seven years, is slowing down and may yet tip into recession. If it does, the world economy will face its worst crisis since the Great Depression of the 1930s.

Mention of that awful decade should make politicians shudder. Then, governments responded to economic crisis by embracing protectionism, with disastrous results. This time should be (and so far is) different. There have been a few unwelcome signals—the odd tariff increase here, introduction of capital controls there—but as yet nothing like a retreat from free trade. Indeed, the strongest bulwark against a return to protectionism is the existence of the WTO itself: the GATT, after all, was set up after the second world war precisely to avoid a repetition of the economic follies of the 1930s.

The danger is that the global economic wobble will prompt politicians to scramble for shelter. Some American and European companies are already being overrun by competition from Asian goods, made much cheaper by drastic devaluations. If a global recession does come to pass and markets start to shrink, more businessmen and workers may demand protection from their governments.

This, therefore, may be an odd time to call for trade to be made freer still. Why rock the boat when the weather is stormy? And is trade not free enough already after half a century of liberalisation?

It is not. The lesson of the 1930s is that free trade, far from being a luxury affordable only when times



are good, becomes utterly essential when times are bad. This survey will argue that the time is ripe for a new round of global trade talks not in spite of but because of the current economic troubles.

The argument that trade is already free enough—or even, as some claim, "unfettered"—is unsustainable. Plenty of barriers to trade remain. Indeed, as conventional trade barriers, such as tariffs and import quotas, have come down over the past half-century, others, such as discriminatory regulations, have sprung up to take their place. Moreover, as the global economy has become more integrated, the meaning of "free trade" has become more elusive. Perhaps 20 years ago most trade negotiators would have settled for the definition in an elementary economics textbook: an absence of tariffs and quotas on goods. This will no longer do. Yet a more up-to-date definition is hard to pin down.

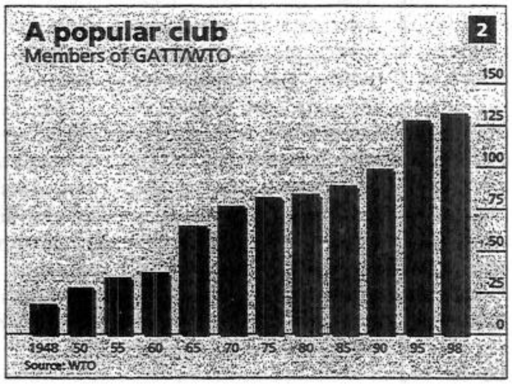
For a start, trade is not restricted to goods. Services too, from telephone calls to architects' plans, can be sold internationally. Instead of straightforward cross-border exchanges, the vehicle for trade in both goods and services can be foreign direct investment. Moreover, economists and trade negotiators realised long ago that barriers to trade are not confined to tariffs and quotas. National regulations can also thwart trade.

The Uruguay round of GATT negotiations reflected these changes in the nature both of trade and of trade barriers. The treaty to seal the round, signed in 1994, enormously increased the scope of trade rules. The WTO not only took over the old GATT, which dealt with trade in goods, but also took on new agreements on services, investment, intellectual property, technical barriers to trade, sanitary measures and plant health. Agriculture and textiles, sectors subject to extraordinary protection, were brought into GATT rules for the first time. The WTO also has a new system for settling trade disputes, with more teeth than that of the GATT. Since the end of the Uruguay round, agreements have been reached on trade in telecommunications, financial services and information technology.

The telecommunications deal is a good example of how trade and trade rules are evolving. It includes a commitment, made by most of the 69 countries that signed the deal, that telecoms regulators will be independent of suppliers. The idea is to prevent national monopolists from being both player and referee, and keeping out foreign competition. In effect, part of domestic competition policy has been subsumed into world trade rules.

However, the transformation of trade rules is incomplete. There are as yet no coherent rules to deal with either foreign direct investment or competition policy. Regional agreements on trade and economic integration have proliferated, prompting questions on whether these help or hinder global free trade. And plenty of trade barriers of the old-fashioned kind remain. Talks on liberalising trade in agriculture and services are already scheduled to begin in the next couple of years. There is easily enough work to justify launching a new round of global trade talks.

At the same time the spread of freer trade, and the success of the WTO, is causing friction between free-traders and defenders of ideas such as environmentalism and human rights. Until recently, the WTO has fought shy of confronting its critics. In preparing for a new round, trade ministers must also tackle these political problems. This survey will offer some suggestions on how they might set about it.



Why trade is good for you

A short tour of economic theory

ECONOMISTS are usually accused of three sins: an inability to agree among themselves; stating the obvious; and giving bad advice. In the field of international trade, they would be right to plead not guilty to all three. If there is one proposition with which virtually all economists agree, it is that free trade is almost always better than protection. Yet the underlying theory is not readily understood by non-economists. And the advice that follows from it—protection does not pay—is seldom wrong.

The argument for free trade is based on the theory of comparative advantage. This is one of the oldest theories in economics, usually ascribed to David Ricardo, an Englishman who wrote in the early 19th century. To see how it works, imagine two countries, East and West, which both produce two kinds of goods, bicycles and wheat. In a year, an Eastern worker can make two bikes or grow four bushels of wheat. A Westerner, however, can manage only one bushel or one bike. Each country has 100 workers, and initially both of their workforces are split evenly between the two industries. So East produces 200 bushels of wheat and 100 bicycles, whereas West produces 50 bushels and 50 bikes (see table 3, first panel).

Since East can produce both wheat and bicycles more cheaply than West, it has an absolute advantage in both industries. Even so, Easterners will benefit from trading with Westerners. This is because East is relatively more efficient at growing wheat, where it is four times as productive as West, than it is at making bikes, where it is only twice as productive. In other words, it has a comparative advantage in wheat. At the same time, West has a comparative advantage in making bikes, even though it has no absolute advantage in anything.

According to Ricardo's theory, both countries will be better off if each specialises in the industry where it has a comparative advantage, and if the two trade with one another. Specialisation increases world output. Suppose that East specialises in wheat growing, shifting ten workers from its bicycle factories to its fields, and producing 240 bushels and 80 bikes. West moves 25 workers from wheat farming into bike making, where its comparative advantage lies, and produces 75 bikes and 25 bushels. Global production rises (see second panel).

The point of economic activity, however, is not to produce but to consume. Both countries can enjoy more bikes and more wheat if they trade on terms at which both will gain. If East is going to import bikes, it will pay no more than two bushels in return (faced with a higher price, it would be better off moving workers back to the bike factory). Similarly, West will pay no more than one bike per bushel. Suppose that the "terms of trade", as economists call the ratio of export to import prices, are set at one-and-a-half bushels per bicycle, and that 33 bushels are traded for 22 bikes. The result (third panel) is that both countries are better off.

Comparative arithmetic 3

Autarky	Wheat	Bicycles
East	200	100
West	50	50
Total	250	150
Specialisation	**Wheat**	**Bicycles**
East	240	80
West	25	75
Total	265	155
Trade	**Wheat**	**Bicycles**
East	207	102
West	58	53
Total	265	155

In essence, the theory of comparative advantage says that it pays countries to trade because they are different. West's relative deficiency in bike manufacture is less than in wheat farming. It is impossible for a country to have no comparative advantage in anything. It may be the least efficient at everything, but it will still have a comparative advantage in the industry in which it is relatively least bad. And even if a country were the most efficient in every industry, giving it an absolute advantage in everything, it could not have a comparative advantage in everything. In some industries, its margin would be more impressive than in others.

Economists' next argument for free trade is that opening up markets to foreign suppliers increases competition. Without free trade, domestic companies may have enjoyed monopolies or oligopolies that enabled them to keep prices well above marginal costs. Trade liberalisation will undermine that market power. Competition should also spur domestic companies to greater efficiency because they will not be able to pass on the costs of slackness in higher prices.

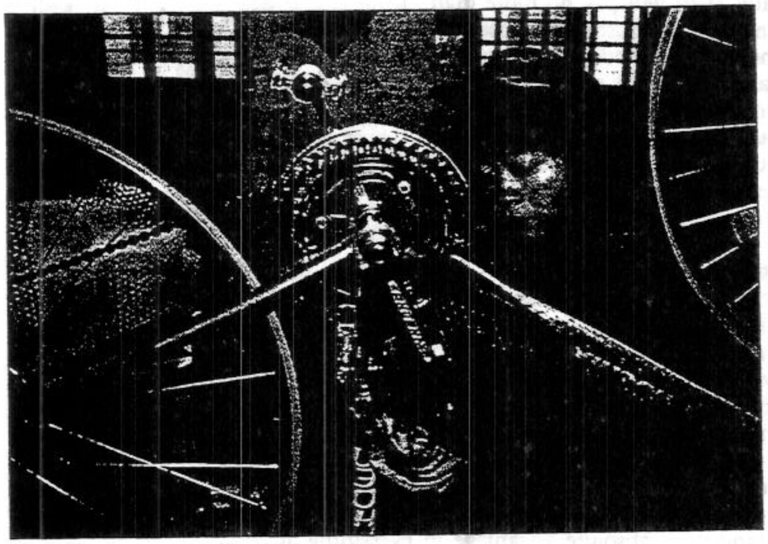
In addition, free trade means that firms are no longer limited by the size of their home country, but

An angelic mix

ECONOMIC theory strips the world of its complications in the hope of making sense of it. Real life is a lot messier, but economic theory can still help to explain it. Take a look, by way of illustration, at the trading activities of one city, Los Angeles, and the forces that have shaped it.

- **Luck.** LA's best-known export is films. Nowadays, this is easy enough to explain in terms of comparative advantage. California's lead in movie making is enormous. Initially, good light and fine weather may have helped, but other places had those too. Luck, a hundred years of practice and economies of scale probably mattered more.
- **Geography.** Thanks to a deep harbour on the Pacific coast, the twin ports of Los Angeles and Long Beach are the busiest in America and the third busiest in the world. Moving containers in and out of the ports and to and from the rest of America is a big business in itself. The Port of Los Angeles reckons that it supports over 250,000 jobs in southern California—more than one in 24 of the region's total.
- **People.** LA's location helps explain its rich ethnic mix. Besides Latinos, whites and blacks, this includes Chinese, Koreans, Vietnamese, Japanese, Indians and many others, creating an intricate pattern of international commerce. Linking up all these different networks can create further business. In El Monte, one of the LA conurbation's 88 municipalities, John Leung of Titan Group, a property company, plans a business park aimed at helping small LA companies find their way into international trade.
- **Technology.** The days when raw materials were produced in one country and turned into finished goods in another are long gone. The making of even the simplest goods is chopped up into a number of different stages, reflecting relative costs in different countries and falling international transport costs. Charlie Woo, chief executive of Megatoys, explains how a child's pinwheel, consisting of plastic sails pinned to a stick, is made in three different countries. The plastic is produced in America and cut to shape in China. The toy is then assembled in Mexico and shipped to LA for distribution.
- **Niches.** Some of LA's textile industry, explains Jack Kyser, a local economist, has gone south to Mexico in search of cheaper labour, but by no means all. Mexico is fine for run-of-the-mill jeans, but suppose that a smart department store quickly needs a new line of high-fashion dresses—perhaps just a few dozen, each selling for a few hundred dollars. A month or two later the style will be dead. The store might buy the dresses from one of the designers exhibiting in tiny spaces on 9th Street.

Many of those displaying here, explains Else Metchek of the California Fashion Association, are based in garages. Plenty more started up in them and moved on to greater things: one, launched 12 years ago, now turns over $85m a year. Even California style, however, has to accept a measure of globalisation. Knitwear production is so labour-intensive that it is moving abroad—not to Mexico, but to China.



Peddling comparative advantage

can sell into bigger markets. In industries where average production costs fall as output increases, producing economies of scale, this means lower costs and prices. In such industries, trade also increases the variety of products on offer. If a car manufacturer, say, were limited to its home market, it would have a choice between producing small quantities of a number of models and large quantities of just a few, which could be produced more cheaply thanks to economies of scale. But given free trade, it would be able to produce more models because they could all be produced in large enough numbers.

All these arguments revolve around re-allocating resources to produce one-off improvements—what economists call the "static" gains from trade. But they think there are also "dynamic" gains to be had: freer trade can mean faster economic growth.

In recent years, theories of economic growth have become much more sophisticated. Although economists have long realised that productivity growth is a prime determinant of general economic growth, models of growth in the past made no attempt to explain productivity growth, but assumed that it was "exogenous" (caused by outside factors). Modern "endogenous" growth theories, however, do try to explain productivity growth.

Freer trade can play a part in this in a number of ways. For one, by making markets bigger it creates more scope for "learning by doing"—ie, for firms to become more efficient with repetition. Larger markets also offer bigger incentives for firms to invest in research and development. Moreover, trade disseminates knowledge and technology. Simply by participating in international markets, countries are exposed to other countries' techniques, and have an incentive to copy and improve on them.

All this can make the relationship between trade, technology and growth quite complicated. For example, freer trade does not necessarily mean faster growth all the time. If a country's comparative advantage lies in slow-growing, traditional industries, it may cut back its production in other, faster-growing industries, so its growth rate may fall. But later on it will benefit from the technical advances of countries with a comparative advantage in faster-growing industries, which will give it better computers, more advanced drugs and so forth than if every country had tried to make everything for itself. Slower growth in the short run might therefore be balanced by faster growth in the long term.

Pinning down the link between freer trade and growth is not easy. One problem is how to measure the openness of a country's trade policy: trade barriers can take many different forms. Even so, there is good reason to believe that freer trade and faster growth generally go together.

Exceptions that prove the rule

So is free trade always the best policy? Not quite. One well-known exception is the "optimal tariff" imposed by countries that are big enough to exert an influence on the world prices of the goods they trade. By raising tariffs, they can significantly reduce world demand, cutting the world price of the good and tilting the terms of trade in their favour.

A second example, which came to prominence in the 1980s, is "strategic trade policy". In an industry with economies of scale, the imposition of a tariff, by reserving the home market for a domestic firm, allows the firm to cut its costs and, with luck, to undercut foreign rivals in overseas markets. Economists and politicians reckon that this might work in civil aircraft, semiconductors and cars.

However, it is rare for governments to be powerful enough to set optimal tariffs—and even rarer for them to have enough information to do so with pinpoint accuracy. They would need to be able to recognise a likely business proposition, and get their maths absolutely right. Even then, rival countries could follow suit, leaving them all worse off. ■

Border battles

Conventional trade barriers are coming down, but not quickly enough

THROUGHOUT eight rounds of GATT talks, tariffs were trade negotiators' staple fare (see table 4, next page). Only in the sixth, the Kennedy round, did the diplomats even begin to add other trade barriers to their diet. After half a century at the table, you would think, they must have made plenty of progress. And so they have; but it is surprisingly hard to say how much, and disappointingly easy to conclude that they still have a lot on their plates.

A commonly quoted statistic is that when the GATT was founded, developed countries' tariffs on manufactured goods averaged around 40%. According to the WTO, by 2000 that average will have tumbled to less than 4%. Most developing countries have further to go, but these figures convey the right message: trade has become freer over the past 50 years. At the same time the figures' precision is misleading.

In particular, it is hard to pin down a meaningful average tariff rate for the late 1940s. Outside the United States, tariffs at that time were not even the main barriers to international commerce. After the first GATT talks in 1947, *The Economist* sniffed that: "For many of the countries involved... customs tariffs are at present without any influence on the volume of trade." Currencies were not freely convertible, and imports were subject to licences and quotas. The trouble with quotas is that, once filled, they completely insulate domestic producers against foreign competition, and bestow monopolies on the importers that hold them. One of GATT's achievements was to establish tariffs as the least bad method of protection, and to negotiate maximum tariff rates (so-called "bound" levels). By now nearly all OECD countries have bound all, or almost all, of their tariffs.

Another reason for treating historical data with caution is that in GATT's early days most tariffs were "specific", ie, based on the volume of imports rather than their value. This, according to a paper by Douglas Irwin of Dartmouth College*, explains why America's average tariff rate looked far lower in the early 1950s (12%) than in the 1930s (50%). Four-fifths of that fall was due to inflation.

However, even today, measuring protection remains difficult. Quotas, licences and specific tariffs are still around. Although WTO members are committed to imposing tariffs on a "most-favoured-nation" (MFN) basis, which is meant to be non-discriminatory, in practice most of them do discriminate. Imports from partners in regional trade agreements are tariff-free. Developed countries levy special low tariffs on some goods from poor countries.

Even the MFN tariffs actually imposed may be far below the "bound" levels. Applied tariffs are therefore a better guide to protection than formal WTO commitments, but there is a risk that they may be increased abruptly. For example, Nigeria's bound tariff on most agricultural goods is 150%, and few industrial tariffs are bound at all; its applied tariffs now average 23%, but they have changed frequently in the past few years.

In addition, "average" tariffs can hide as much as they reveal. The measure used by the WTO is to weight tariffs by the share of goods in a country's imports. But if a tariff is pitched so high that it chokes trade altogether, its weight becomes zero, which is clearly nonsense.

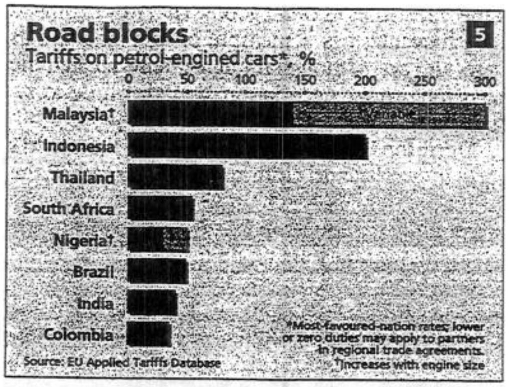
The OECD's annual survey of trade barriers† uses several different weighting systems, none of them perfect. Under one, tariffs are weighted by the shares of goods in the domestic production of the importing country, but this gives too much weight to highly protected products. In another, the simple average of all tariffs is used unweighted, which means it takes no account of the relative importance of different goods. In a third, the import weights used are OECD averages rather than the country's own. Yet OECD countries often protect the same industries, depressing their weights across the board. Even among developed countries, which tend to have lower tariffs than developing ones, the use of different weights can yield markedly different results. In 1996, estimates the OECD, Canada's import-weighted average tariff was 5.7%, its simple average was 9.2% and its production-weighted average was 12.1%.

What GATT did 4

GATT rounds

Date	Venue/name	Subject
1947	Geneva	Tariffs
1949	Annecy	Tariffs
1951	Torquay	Tariffs
1956	Geneva	Tariffs
1960-61	Geneva (Dillon round)	Tariffs
1964-67	Geneva (Kennedy round)	Tariffs and anti-dumping measures
1973-79	Geneva (Tokyo round)	Tariffs, non-tariff measures, "framework" agreements
1986-94	Geneva (Uruguay round)	Tariffs, non-tariff measures, rules, services, intellectual property, dispute settlement, textiles, agriculture, creation of WTO, etc.

Source: WTO



*Most-favoured-nation rates; lower or zero duties may apply to partners in regional trade agreements. †Increases with engine size

Source: EU Applied Tariffs Database

No matter which weighting system is used, averages tell you nothing about extremes. In manufacturing trade high tariff rates, or spikes, are less common than they used to be, but even in developed countries they still persist. For example, trucks attract a tariff of 25% in America and 22% in the European Union. High tariffs in the food industry often reflect heavy protection for agriculture. America's tariff on peanut butter is 132%; the EU slaps anything between 46% and 215% on fruit juices.

Developing countries can be spikier still. Malaysia's average import-weighted applied tariff fell from 15.2% in 1993 to 8.1% last year; well done. But tariffs in agricultural goods and car manufacture can go up to 145%, on top of licensing restrictions. Many developing countries make heavy use of high tariffs to protect their car makers (see chart 5).

And so to work

All of this suggests that the WTO still has plenty of tariff-cutting to do. The best place to start would be manufactured goods. In developed countries, the tariffs on some of them are so low that they are barely worth collecting. They should be scrapped. In developing countries, average levels still have a long way to go—Mercosur's common external tariffs, for example, range up to around 20%. And there are plenty of spikes that need blunting.

The WTO also has to sort out textiles and agriculture, two of the world's most protected industries. The Uruguay round brought both of these under the

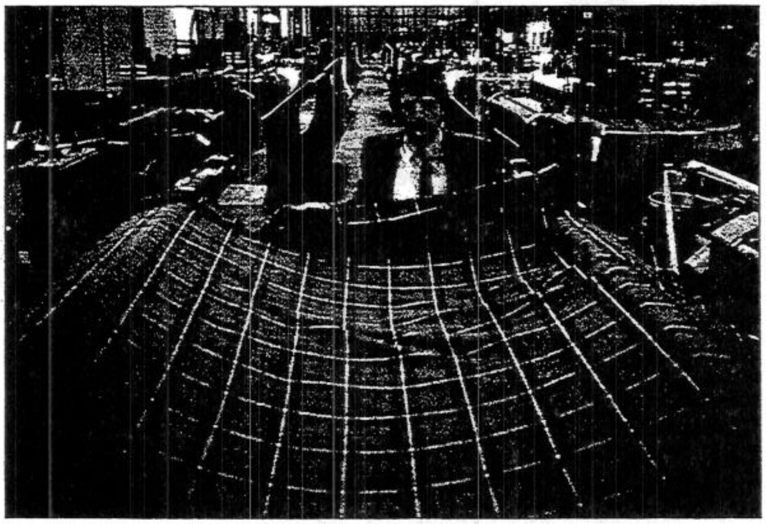
---

* "Changes in U.S. Tariffs: The Role of Import Prices and Commercial Policies". *American Economic Review*, September 1998

† "Indicators of Tariff and Non-Tariff Trade Barriers". 1997 edition

same trade rules as other goods. That this counts as an achievement is a measure of how bad things were, and are. Although the agreements created the basis for further negotiation, they did little to reduce protection.

Until the Uruguay round, much of the world's trade in textiles and clothing was governed by the Multifibre Arrangement (MFA), a cat's cradle of bilateral import quotas. Under the new textiles agreement, these quotas are being increased and items covered by the MFA are gradually being brought under GATT rules. "Gradually" is the operative word. At the start of 1995, WTO members shifted items accounting for 16% of their 1990 imports into the new system. Another 17% followed at the beginning of this year. A further 18% is due to move over in 2002, and the rest follows in 2005.



Feel the width of those textile quotas

To slow things down further, importing countries (mostly developed ones) chose to start the transition from the MFA to the GATT with items on which quotas were not fully used. In the early years, therefore, it makes very little difference. And even when all the quotas have gone, trade will be no more liberal than at the outset, because high tariffs will remain. The United States, for instance, levies tariffs of between 14% and 32% on most synthetic, woollen and cotton clothing. A $25 pair of leather shoes imported into Japan attracts a tariff of 160%. No wonder that developing countries, which tend to be textile exporters, feel short-changed—even though their own markets are rarely monuments to open trade.

In much the same way as the textiles agreement, the Uruguay round's agreement on agriculture brought farm trade under GATT disciplines for the first time. In a useful recent study*, Timothy Josling, an economist at Stanford University, lucidly explains the fiendishly complicated results. The agreement did three things. First, with a few temporary exceptions, it converted all non-tariff barriers and unbound tariffs into bound tariffs. These tariffs have to be cut by an unweighted average of 36% between 1995 and 2000. Second, it prohibited new export subsidies and cut existing ones. And third, it began to tackle the domestic subsidies which, in effect, protect farmers against foreign competition in much the same way as tariffs.

This provides the starting point for new negotiations, due to begin next year. There is plenty to talk about. Some of the tariffs that have taken over from quotas are clearly set at a level designed to stop trade altogether: 300% on butter in Canada, 550% on rice in Japan, 215% on frozen beef in the EU, 179% on sweet powdered milk in America.

Further liberalisation, argues Mr Josling, is urgently needed, not only because tariffs and subsidies remain high but because the countries trying to join the WTO include big agricultural producers with large state-owned enterprises, such as China and Russia. There are some promising signs. In its recent Farm Act, America took another step towards uncoupling support for farmers from production subsidies. American farmers, keen to export, have been lobbying for a renewal of fast-track negotiating authority. In the EU, meanwhile, the prospect of enlargement to the east is increasing the pressure for reform of the common agricultural policy.

Then again, the difficulties should not be underestimated. If farm prices were to tumble, both America and Europe might be loth to relinquish export subsidies, which in recent years have been kept down by firm markets. America's sugar, dairy and peanut farmers have so far held on tenaciously to protection, and will try to go on doing so. Europe, too, can still muster plenty of opposition to free farm trade. Agriculture nearly scuppered the Uruguay round altogether. It seems more than capable of disrupting the next one. ■

---

* "Agricultural Trade Policy: Completing the Reform". Institute for International Economics, Washington, DC, April 1998

Trade by any other name

Does the WTO need special rules for foreign direct investment?

THE growth of international trade is not the only, or even the most impressive, measure of recent global economic integration. Between 1990 and 1997 the value of goods crossing international borders grew by just under 60% in dollar terms, whereas foreign direct investment (FDI) over the same period nearly doubled. Most of this investment went from one OECD country to another, but a growing share is now going to developing countries, mainly in Asia (see chart 6, next page).

Time was when economists considered FDI as a substitute for trade. Building factories in foreign countries was one way of jumping tariff barriers, which helps to explain the presence of foreign car manufacturers in countries from Brazil to South-East Asia. But these days economists are just as likely to think of FDI and trade as complementary.

For example, a firm can use a factory in one country to supply neighbouring markets. Ring Gateway 2000, an American computer company, from anywhere in Europe, and your call will be answered (and your order assembled) in premises near Dublin airport. A lot of FDI is also explained by the trend towards splitting up manufacturing among several countries. America's official statisticians estimate that one-third of all their country's trade takes place within companies. And some investments, especially in service industries, are essential prerequisites for selling to foreigners. Who would buy a Big Mac in London if it had to be sent from New York?

All this raises two questions. First, why should governments treat FDI, and the trade associated with it, any differently from other kinds of trade? Answer: in principle, they should not—although in practice the issue is hotly disputed. Second, what is the equivalent for FDI of free trade? The answer to that is more complicated—and just as hot politically.

Put out the welcome mat

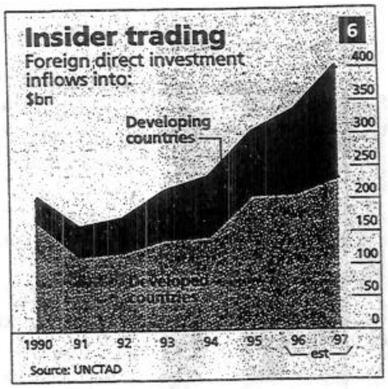
Edward Graham, an economist at the Institute for International Economics, a Washington think-tank, suggests some criteria, similar to existing WTO rules on trade in goods. One is "national treatment". In the GATT, this means that imports, having crossed the border, should not be subject to special taxes or other forms of discrimination. For FDI, however, there could be a weak or a strong version of national treatment. Under the weak one, foreign investors, once established, would be treated the same as locally owned companies. Under the strong version, foreigners would also have the same right of establishment as locals. This is the FDI equivalent of zero tariffs.

In line with another GATT doctrine, says Mr Graham, foreign investors should have "most-favoured-nation" status: host governments should not discriminate among investors from different countries. There should also be a forum for settling disputes. If the host government were to default on its commitments, investors should be able to seek redress before special tribunals or national courts. This would be a change from the WTO's existing procedures, in which the only plaintiffs and defendants are governments.

Existing WTO rules on investment give foreign investors far fewer rights than these. The current rules were laid down in three separate Uruguay round agreements. The first, the Agreement on Trade-Related Investment Measures (TRIMS), deals only with investment in manufacturing. It stops governments from requiring investors to use a minimum level of local inputs, and to export at least as much as they import.

The second, the General Agreement on Trade in Services (GATS), recognises "commercial presence" (WTO-speak for setting up shop) as one of the ways in which services are traded, but it does not oblige WTO members to throw open all their service industries to all comers. Although they must subscribe to its general principles, they need open up only those industries listed in their "schedules of commitments". In many cases, these lists are pretty short.

The third, the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS), obliges WTO members to grant minimum periods of protection to copyrights, patents and trade marks. Developing countries have more time than rich ones to bring their laws into line. The TRIPS agreement protects foreign investors by stopping locals from pinching their inventions and brands.

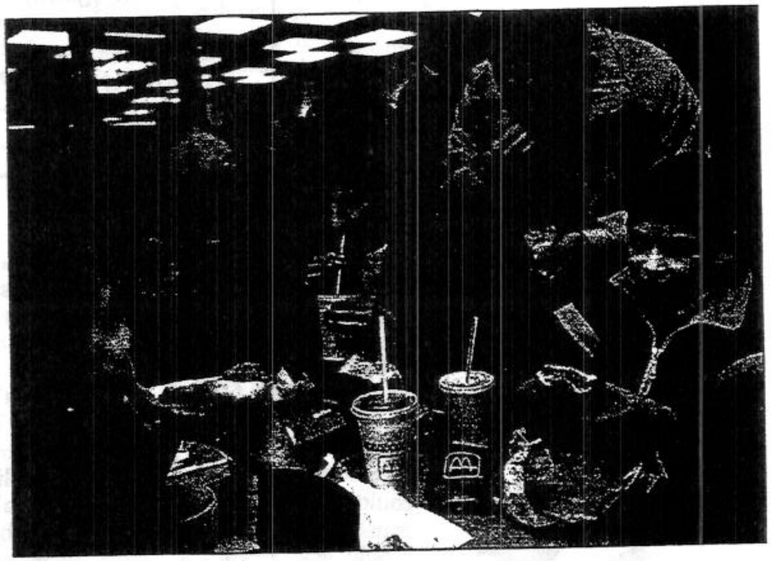


The chances of creating a single FDI accord incorporating Mr Graham's criteria look slim. True, the principles he outlines are already enshrined in regional trade agreements and some of the 1,600 bilateral investment treaties (BITS) signed so far. Companies established in one EU country are free to go wherever they like within the Union. The North American Free-Trade Agreement allows investors to take host governments to tribunals, but the few cases that have been brought so far have proved controversial.

Moreover, the OECD's efforts to build a multilateral investment agreement (MAI), seen by some as a springboard for a WTO agreement, stalled earlier this year. Governments have been unable to agree on how far investment should be liberalised. Environmentalists, trade unions and other non-governmental organisations denounced it as a multinationals' charter that would leave governments powerless to protect either workers or the planet. That scared politicians into retreat.

The OECD deal, on which talks are due to resume soon, might not be worth much anyway. Most OECD countries already have liberal FDI regimes, and the MAI would not have addressed their biggest sin: the huge subsidies they shell out to attract foreign factories. Even so, the MAI's troubles do not augur well for a WTO agreement.

On top of this, many developing countries are opposed to any FDI deal at the WTO. Manmohan Singh, the finance minister who began India's economic liberalisation in the early 1990s, explains that several BITS have been signed granting national treatment to foreign investors, but that: "We are not ready as yet for right of establishment. You have to remember our history as a colony. The East India Company came here as a trader and ended



Making FDI palatable

Brittan disagrees

up owning the country."

So what should governments do? America still favours persevering with the OECD talks. "It's a little premature to say it won't produce anything. It may produce a lot," says Charlene Barshefsky, the United States trade representative. But Sir Leon Brittan, the EU's trade commissioner, who has long been pressing for a WTO accord, disagrees. "The difficulties of the MAI should make the climate more propitious," he says. "Among developed countries, there's been a feeling that although we'd like to talk at the WTO, let's not get in the way of the MAI." By way of reassurance to developing countries, Sir Leon stresses that he is not proposing an investment free-for-all.

Given the way trade negotiations work, a WTO deal of some sort is not impossible. India's reluctance might be overcome if, say, rich countries moved smartly to open their textile markets. And a free-for-all could be avoided by an agreement similar to the GATS, in which governments settled on general principles but opened up sector-by-sector.

The larger question is whether a WTO agreement on FDI is worth pursuing. These days, governments that were once suspicious of foreign investors are welcoming them with open arms, especially in manufacturing. They hope that investors will bring expertise and technology that will be passed on to local firms and workers, helping to sharpen up their whole economy. Foreign investors, for their part, have needed no prompting to pour money into factories in developing countries, especially China. And where FDI is not forthcoming, as in sub-Saharan Africa, investors are unlikely to be swayed by the reassurance offered by a WTO accord.

Open wider

However, argues Bernard Hoekman, an economist at the World Bank, there are still important barriers to foreign investment in service industries. He reckons that under the GATS, rich countries agreed to open about half of the service industries covered by the agreement, but developing countries managed only one-sixth—and in many of these markets the opening was less than complete.

In failing to liberalise services, governments are harming their own economies. Services such as transport and telecommunications are vital for the efficient functioning of the rest of the economy. Indeed, write Mr Hoekman and Carlos Primo Braga, another World Bank expert, in a recent paper*, protecting service industries can undo the benefits of liberalising goods trade. Cutting tariffs will expose manufacturing industries to foreign competition, but if poor roads, railways and other communications stop goods from getting to market quickly, those manufacturers will be hobbled.

Just look, say the two authors, at the gains Chile reaped from liberalising its merchant shipping industry. Increased competition spurred the introduction of new technology, such as refrigeration, which helped Chile to become a successful exporter of salmon and kiwi fruit. By contrast, the United States is paying dearly for the Jones Act, which restricts shipping between American ports to American-flagged ships. This has pushed freight rates to between twice and four times what they would be under free trade.

Moreover, foreign investment in services may bring more skills into host economies than investment in manufacturing, because goods production is more easily split into its constituent parts than services production. A manufacturer might set up a "screwdriver" plant in Mexico while keeping its research and development department in America and its high-tech component factory in Europe. But a company running hotels or selling banking, insurance or telecoms services will have to train local hoteliers, bankers, insurers and telecoms engineers.

That advantage may not last. In time, telecommunications technology will develop to the point where more and more services can be traded across borders. There may no longer be a need for, say, an insurer to have an office in Caracas in order to sell to the Venezuelan in the street. The know-how can stay in New York or Zurich. That prospect should spur developing-country governments into action: liberalise now, or lose out.

If Mr Hoekman is right, then the WTO's best course might be to concentrate on further liberalisation within the GATS, rather than strive for a new across-the-board deal on FDI. Still, strengthening safeguards for investors in goods industries would be a welcome bonus.

---

* "Trade and Protection in Services: A Survey". *Open Economies Review*, 1997

Commerce and contestability

The marriage of trade and trustbusting

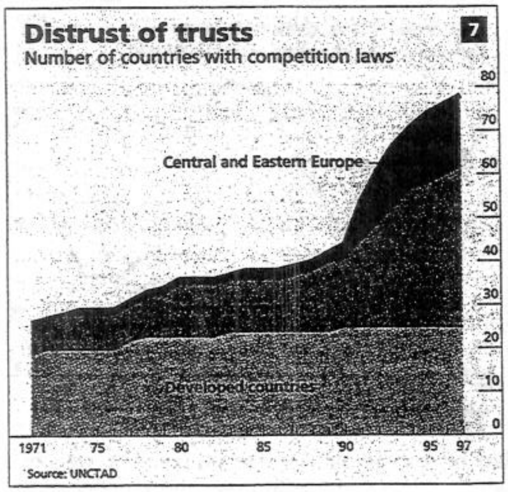
SUSPENDING disbelief, suppose that the world's governments had abolished all barriers to international trade, and allowed foreign investors to set up businesses on the same terms as their own people. Would free trade have arrived?

Probably not. Some markets would still not be open to competition, whether from home or abroad, either because as natural monopolies they could not support more than one supplier, or because entrenched monopolies and oligopolies would keep competitors out. Private barriers to trade might replace those put up by the state.

This is why in many countries the liberalisation of trade and FDI policies has gone hand-in-hand with the establishment of competition authorities. Since 1990, some 30 developing countries and former communist countries have introduced antitrust laws. In all, nearly 80 countries now have such laws (see chart 7, next page).

Yet the actions of a country's competition authority can affect foreign companies' ability to enter its market, or spill over into other countries' markets. Some economists and politicians feel that competition policies need to be more closely coordinated internationally; some, indeed, argue that since trade and competition policies have become so intertwined, the WTO should embrace competition policy as well.

For



Distrust of trusts 7

Number of countries with competition laws

Source: UNCTAD

Competition policies can spill over into trade in several ways. Examples include:

- **Aircraft.** A merger or alliance between two companies from the same country might affect competition in other countries. Thus American regulators were happy for Boeing, a giant manufacturer of civil aircraft, to take over McDonnell Douglas, an ailing smaller firm. But the EU's competition commissioner, Karel Van Miert, was worried that the merger would reduce competition in Europe, to the detriment of Boeing's European rival, Airbus, though in the end the deal went through.
- **Airlines.** A merger or an alliance between two companies from different countries might affect competition in the two markets in different ways. For example, American regulators nodded through the mooted alliance between British Airways and American Airlines because they felt it was unlikely to hurt competition at the American end of transatlantic routes. But Mr Van Miert, looking at the deal from a European angle, thought it would do some damage, and insisted that BA give up some take-off and landing slots at London's Heathrow airport.
- **Fortress Japan, as seen from Detroit.** A monopoly or group of companies in one market may be helped by slack competition policy to block entry by foreigners. Time and again in the past two decades, American businessmen have complained to the United States trade representative that the Japanese market was impossible to break into, despite low tariffs. In 1995, America came close to imposing punitive tariffs on Japanese luxury cars in retaliation against the exclusive ties between Japanese car makers and showrooms. It eventually backed down. In 1997, America lost a case at the WTO in which it had argued that Fuji, Japan's market leader in photographic film and paper, was keeping Kodak, America's number one, out of the Japanese market by controlling distribution and retailing.
- **International cartels.** Companies may collude to fix prices internationally, or agree not to compete in each other's markets. To fight this, several competition authorities may need to exchange information and act jointly.
- **Export cartels.** In some countries companies are allowed to form export cartels because the resulting higher prices will affect only foreign customers. In practice such cartels account for only a small proportion of trade, but there is a strong economic ar-

The Wired Trade Organisation

THROUGHOUT history, international trade has been helped by all manner of technical advances, from the development of the tea clipper to the invention of powered flight and the telecommunications revolution. The last of these, and especially the Internet, could have a huge effect on trade in the next few years.

Electronic commerce should boost trade in goods: hard-to-find books or music recordings, for example, have become easier to track down on the Internet. But trade in services should benefit even more. Anything that can be put into digital form will be tradable. For example, architects will be able to send and amend designs electronically. Doctors may be able to diagnose and dispense to patients abroad whom they never meet.

One of the best things about electronic commerce is that it is fairly free from interference by governments. America wants to keep it that way. It has called for a WTO accord on electronic commerce that would keep the electronic transmission of digitised information free of customs duties.

This looks like a trade diplomat's dream. In the absence of any barriers at present, nothing needs to be negotiated away. The WTO's agreement on trade in telecommunications services already guarantees the freedom of some aspects of electronic commerce. And a deal can almost certainly be done without fast-track negotiating authority for President Clinton.

Can you drop it on your foot?

Despite all these plus points, electronic commerce raises two difficulties for the WTO. First, it blurs the distinction between a good and a service. This matters because WTO rules treat goods and services differently. Goods tend to be subject to tariffs; services are not, but trade in services is limited by restrictions on "national treatment" or quantitative controls on access to foreign markets. So the rules that will be devised for electronic commerce may affect the choice between physical and digital methods of trade.

For example, a compact disc sent from one country to another is clearly a good, and will incur a tariff as it crosses the border. But if the music on the disc is sent electronically from a computer in one country to a computer in another, is it still a good, even though it can no longer be dropped on your foot? Customised data and software, which can also be put on CD, are usually treated as services. Who can tell?

Second, electronic commerce poses a headache for national regulators, especially in service industries such as medicine and financial advice where suppliers are much better informed than their customers. Watchdogs may be helpless to stop the electronic sale of quack treatments or dodgy investment schemes.

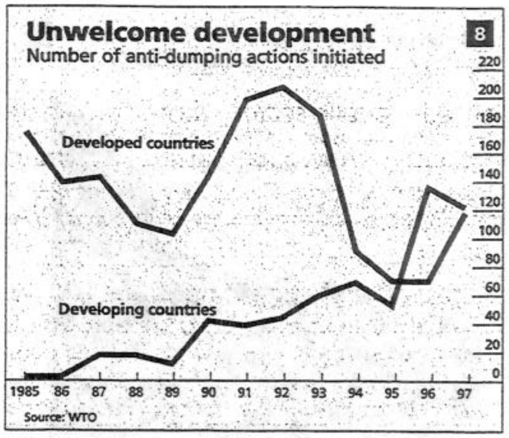
They could impose trade restrictions, insisting, for example, that financial firms selling on the Internet to residents of their country must also have an office there; or they could work more closely together, with officials in the seller's country monitoring cross-border sales on behalf of regulators in the buyer's country. But such invigilation would take away some of the Internet's free-wheeling charms. More sensibly, they might decide to leave well alone and let the buyer beware.

gument for outlawing them all.

The WTO is divided over how to deal with these problems. The EU is keen on a WTO agreement on competition policy. Its trade commissioner, Sir Leon Brittan, would like all the organisation's members to commit themselves to introducing effective competition laws, covering mergers, monopolies and restrictive business practices. These laws would have to meet minimum standards: proceedings would have to be transparent, and the rules would have to be backed up by sanctions. Domestic and foreign firms would have to have equal access to national courts.

By contrast, Joel Klein, the head of the antitrust division of America's Department of Justice, sees no value in a WTO agreement setting minimum standards. "I strongly suspect that nearly all of the world's 70-odd competition laws, accounting for nearly all important trading nations, would likely meet the requirements of any minimum substantive rules the WTO could adopt," he says. "So what is the point?"

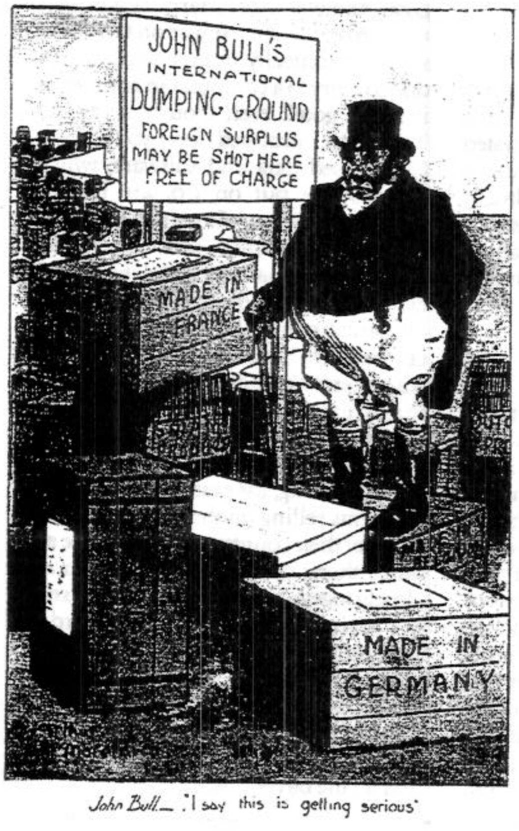
The Americans also say that it would be very hard to agree on competition rules. In the Boeing case, for example, America and the EU read the same facts in different ways. The effects on the two markets were not all that different, because both Airbus and Boeing sell to both American and European airlines. Moreover, Americans and Europeans treat "vertical restraints", such as exclusive deals between retailers and manufacturers, in quite different ways. In the EU most such arrangements are illegal unless they have specific exemptions. In America each case is considered on its merits.


Unwelcome development 8

Number of anti-dumping actions initiated

Source: WTO

Mr Klein thinks there is more to be gained from bilateral arrangements between national competition authorities than from a WTO deal. In particular, he favours greater use of "positive comity", which allows a competition watchdog in one country to ask its counterpart in another to investigate anti-competitive practices on its behalf. The EU and America have had such a deal since 1991.



John Bull— "I say this is getting serious"

Yet when Mr Klein says there is "no point" in a WTO agreement setting minimum standards, he conveniently overlooks the fact that some 60 WTO members still have no competition policies at all. For those countries, some standards, however modest by Mr Klein's lights, would be better than none. And in the past GATT commitments have been treated not as an end in themselves, but simply as the starting point for further improvement. Just as GATT has brought down tariffs in the past 50 years, so the WTO could raise the standards of competition policy.

Whisper it not

However, there is one important link between trade and competition policies that neither America nor the EU wants to discuss. The unspeakable practice is anti-dumping. Under WTO rules, countries are allowed to impose anti-dumping duties on imports sold at below "normal value"—which can mean the price in their home market or the cost of production. The fact that the WTO permits anti-dumping may make it sound respectable. It very rarely is.

Its defenders say that it helps firms deal with "unfair" foreign competition, but in reality it is an arbitrary form of protection. The information from which "normal value" is calculated is often sketchy. Trade authorities listen to producers hurt by "dumping", but take no account of the views of consumers, who might well prefer cheaper goods. In a recent study* of anti-dumping petitions in America, the EU and Canada in the 1980s, Robert Willig, an economist at Princeton University, set out to establish whether the practice was justified on competition-policy grounds, and concluded that in more than 90% of successful petitions it was not.

In time, as developing countries make increased use of anti-dumping duties (see chart 8), the EU and America may become more inclined to talk about the subject. And if more countries adopt competition policies, the rationale for anti-dumping, already weak, will look flimsier still.

---

* "Competition Policy and Anti-Dumping". *Brookings Trade Policy Forum*, forthcoming

Alphabetti spaghetti

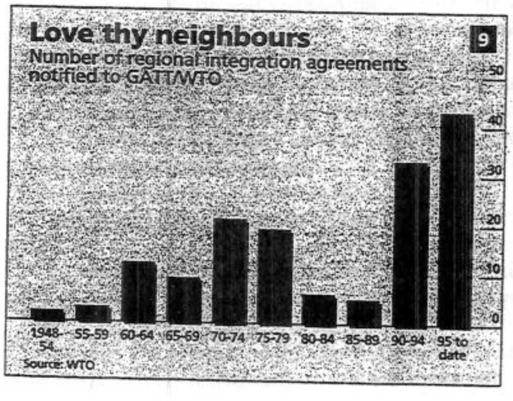
Are regional trade agreements a good idea?

DRIVING around the ever-spreading factories of Tijuana, Enrique Mier y Teran points to one foreign-owned plant after another: JVC, Sony, Messer Griesheim, all making goods for the American market. Then he stops between two factories. The one on the right, he explains, can export TVs duty-free to the United States because it uses tubes made in the one on the left. The one on the left is a product of the North American Free-Trade Agreement's "rules of origin": for a TV to enter America duty-free, some critical components—in this case, the tubes—must be made in NAFTA countries.

Mr Mier y Teran's tour neatly captures economists' ambivalence over regional trade agreements. Thanks to free trade between Mexico and the United States, Americans now import from their southern neighbours things that they used to make for themselves at greater cost. Economists call this "trade creation", and it is all to the good. Yet were it not for the rules of origin, the tube factory would have been somewhere else, probably in Asia. America imports those TVs, or at least the tubes inside them, from Mexico only because NAFTA discriminates against non-NAFTA goods. This is called "trade diversion", and it is not a good thing at all.

In the past few years, economists have had a great many regional agreements to scratch their heads over. Besides NAFTA, the Americas have seen the creation of Mercosur in the south, and one day a Free-Trade Area of the Americas may be born. The European Union, the grandmother of them all, has formed the European Economic Area with the remaining members of the European Free-Trade Association. It also has Europe Agreements with Central Europeans as well as deals with North African countries and Mercosur. In Asia there is APEC, the Asia-Pacific Economic Co-operation forum, which has a grand goal of "free trade in the Pacific by 2020" (but no formal trade deal at present), as well as the ASEAN free-trade area and ANZCERTA, a pact between Australia and New Zealand. All told, in the past 50 years 153 regional trade agreements have been notified to the GATT or the WTO, of which most are still in force. Almost half of these—some, admittedly, revisions of previous deals—have been set up since 1990 (see chart 9).

This deluge of alphabet soup has deeply divided economists. Even language has become an issue. Jagdish Bhagwati of Columbia University, perhaps their most vocal critic, insists on calling them "preferential" rather than "regional" or (heaven forbid) "free" trade agreements, lest anyone forget their trade-diverting effects.



Love thy neighbours 9

Number of regional integration agreements notified to GATT/WTO

Source: WTO

As international commerce has become more intricate, so have the arguments. For a start, regional deals do not take place in a vacuum; they affect the climate of global trade negotiations. But are they, in Mr Bhagwati's phrase, "building blocks" or "stumbling blocks"? The first, says Fred Bergsten of the Institute for International Economics. In the early 1980s, he recalls, the thought that America might embrace regionalism, previously anathema in Washington, shocked the EU into GATT talks. The same thing, he believes, happened in the 1990s: the rise of APEC persuaded the EU to bring the Uruguay round to a close.

The genuine NAFTA article

Mr Bhagwati agrees with Mr Bergsten's reading of the 1980s, but considers his view of the 1990s "fanciful". Politically, he thinks, NAFTA has made the liberalisation of trade between America and poorer countries more difficult. Contrast that with the Uruguay round, in which there was only mild political resistance to freeing America's trade with much poorer India.

A second complication is that manufacturers increasingly carry out different stages of production in different countries. It is this which gives rise to rules of origin. Regional agreements need to spell out whether, for example, a TV assembled in Mexico, which might contain components from all over the world, is Mexican enough to be allowed duty-free into the United States. This is a nonsense, says Mr Bhagwati. Preferential trade ageements create a "spaghetti bowl" of overlapping rules of origin that are sure to distort some investment decisions.

Third, as Bernard Hoekman of the World Bank describes in a recent paper*, regional agreements are pushing the definition of free trade beyond the

---

* "Integrating Domestic Policies". Unpublished

"national treatment" principle. Under national treatment, all goods and services, once inside the country, are subject to the same rules, regulations and taxes, regardless of their origin. Thus all companies are subject to the same antitrust and merger rules. Imported pharmaceuticals, food and electrical equipment, say, have to pass the same safety tests as those produced at home. Doctors and accountants must hold the country's professional qualifications to be able to practise.

Secret weapon

However, such rules can act as barriers to trade. Sometimes they are deliberately used in that way: for example, Mexico used to insist that imported American pigs be vaccinated against hog cholera, a disease that disappeared from the United States in the 1970s. But even if national rules are applied in all innocence, they can still gum up trade. Having been tested for safety at home, goods must be tested again before being sold abroad. Professionals have to take more exams before they can practise abroad. Some countries are now trying to get round these problems by harmonisation or mutual recognition of standards.

The prime examples of such agreements are to be found within the EU. In 1979, the European Court struck down a German ban on Cassis de Dijon, a French liqueur, imposed because of its low alcohol level. This established the principle of mutual recognition: that goods sold in one member state could be sold in all. As well as mutual recognition, in some fields the EU is pursuing harmonisation. For instance, products are subject to common minimum safety standards. National competition laws are supplemented by pan-European legislation. And a dramatic act of harmonisation is due to begin next year, when 11 countries will adopt the same currency.

Within other regional trade agreements, governments have also sought mutual recognition and harmonisation agreements, if not to the same extent as within the EU. Nevertheless, Mr Hoekman points out, a trade agreement is not a prerequisite for a deal on standards. The EU and America have recently struck such a deal even though there is no prospect of a transatlantic trade agreement. In several sectors, ranging from medical and electrical equipment to pharmaceuticals, the two sides have agreed that European organisations can test goods to American standards on behalf of American testing organisations, and vice versa. This falls short of mutual recognition, but is a step in that direction.

This kind of integration is often as much about politics as about commerce. Even so, harmonisation and recognition agreements can both create and divert trade. The new agreement between America and Europe should boost transatlantic trade in electrical goods, but the deal also gives American and European products an edge in each other's markets over goods from other countries. It is worth noting, says Mr Hoekman, that the EU is loth to negotiate mutual-recognition agreements with developing countries.

The WTO does have rules on regional agreements, but they are vague and out of date. Article XXIV of the GATT says that a regional deal must cover "substantially all trade" among its members. Define that if you can. It adds that barriers to non-members must not be raised. This is a weak test of trade diversion.

In 1996 the WTO set up a committee to examine regional trade agreements. It replaced a bunch of working parties on the compatibility of individual deals with Article XXIV. They rarely reached conclusions. Whether the new committee will be more incisive remains to be seen, but it seems unlikely to offer strong criticism of any regional group—if only because of the WTO's 132 members, all but three (Japan, Hong Kong and South Korea) are members of some regional trade club.

The trouble with regional deals is that they discriminate between their members and the rest of the world. The best thing the WTO can do to minimise the damage is not to try to unpick them, but to press on with multilateral trade liberalisation. The more progress it can make on that front, the less room there will be for discrimination.

Turtle wars

Greenery and globalisation do not mix

CESAR LUNA leafs through photographs of a "border reality trip" to Tijuana. Corroding drums of chemicals litter an industrial site open to the elements. There is nothing to keep children out, says Mr Luna, a campaigner with the Environmental Health Coalition in San Diego. Water runs down an escarpment towards people's homes below. He talks about heavy metals and acids contaminating the environment and damaging people's health: "We're not talking about skin rashes and cold sores, but serious long-term health problems."

Mr Luna explains that he is not against a "just" free-trade agreement between America and Mexico, but that in Tijuana some companies are getting away with far too much. The side accords to the North American Free-Trade Agreement, which oblige NAFTA countries to uphold their own labour and environmental laws, have no teeth.

Environmental campaigners such as Mr Luna are becoming more vocal in arguing that world trade is already too free for the planet's good. And if they are cross about NAFTA, they are hopping mad about the WTO.

They have two main complaints. First, a string of national laws aimed at protecting the environment or human health have fallen foul of the WTO's dispute-settlement panels, beginning in 1995 with the very first case to be heard by one of these tribunals. The way in which America administered its clean-petrol laws was held to discriminate against imports from Venezuela.

Since then, say the greens, things have got worse, citing two cases decided earlier this year. In the first, the WTO upheld a complaint by India, Malaysia, Pakistan and Thailand against an American law banning the sale in the United States of shrimps caught

Do not mock turtles

in nets that do not allow sea turtles to escape. Shrimp nets are the biggest killers of sea turtles, says Peter Fugazzotto of Earth Island Institute, a marine-conservation group in San Francisco. It was Earth Island that prodded the American government into taking up the turtle case. The second decision was that a long-standing EU ban on beef treated with growth hormones—ie, most American meat—was incompatible with WTO rules.

The environmentalists also fret that WTO rules may undermine some international agreements to protect the global environment (known as multilateral environmental agreements, or MEAs). Of the total of around 200 such MEAs, about 20 either ban trade in certain pollutants or in goods made using them, or permit the use of trade sanctions as tools of enforcement. For example, the Montreal Protocol bans trade in chlorofluorocarbons (CFCs), which deplete the ozone layer. This trade ban, says Duncan Brack of the Royal Institute of International Affairs in London, played a big part in persuading countries to sign the protocol and end CFC production.

Luckily, so far no government has challenged the trade provisions of any MEA, but that luck may run out. For instance, a country that has not signed the Basle Convention on hazardous wastes (which bans trade in some waste chemicals and metals) might want to import blacklisted goods in order to recycle them or use them in some manufacturing process. Its goods might be barred by a signatory to the convention. The non-signatory could take the reluctant importer to the WTO, and stand a good chance of winning.

All this might worry environmentalists less if the WTO's dispute-settlement system were as limp as the old GATT's, but in fact it is remarkably effective. If one government believes that another is blocking its imports in breach of WTO rules, it can ask for talks. If these fail to resolve the dispute, the complaining government can ask for a panel of trade officials to adjudicate. If the panel finds the rules have been broken, the "guilty" party is supposed to amend its laws or practice to conform with WTO rules. Appeals are possible, but once a final decision is reached, it can be blocked only by a consensus of WTO members. This is a big change from the old GATT system, under which every member (including guilty parties) had the right of veto. So far, no one has ignored a panel decision, because no one wants to jeopardise the credibility of the system of rule-based trade. But if someone did, the offended party could eventually retaliate with trade sanctions of its own.

The WTO's rules do in fact allow countries to impose trade restrictions for environmental and health reasons. Article XX of the GATT permits trade measures "necessary to protect human, animal or plant life or health ... [or] relating to the conservation of exhaustible natural resources if such measures are made effective in conjunction with restrictions on domestic production or consumption."

However, some supposedly "green" rules are not necessary to protect health or the environment at all, but are protectionist in either intent or effect. America's ban on Venezuelan petrol is one example. The EU's refusal to accept beef treated with hormones is another. There is little scientific evidence to suggest that such meat is any more dangerous than hormone-free beef.

Some greens insist that the WTO has no business sitting in judgment on laws passed by sovereign governments. Yet judging the compatibility of national laws with trade accords is precisely what the WTO's dispute-settlement process—set up with the agreement of all its member states—is there for. Both the disputed cases are examples of governments doing more than they should. Americans can surely make up their own minds on which shrimp to eat, helped if necessary by clear labelling. And if Europeans are happy to eat hormone-treated beef, and science judges it safe, why shouldn't they? John Jackson of Georgetown University Law School, an authority on WTO law, says the hormone case is particularly interesting because: "It's the first to grapple heavily with the question of what scientific evidence a government has to present and the degree of risk [to consumers] it is obliged to negotiate." WTO rules suggest that governments can decide how much risk they will accept in imported food. In effect, says Mr Jackson, the EU decided to run no risk at all. But if governments have so much leeway, they could get away with any import ban, however flimsy its scientific foundations.

A shrimp's a shrimp for a' that

A more important reason why the WTO and the greens are at loggerheads is another GATT rule requiring "like products" to be treated identically. This means that with very few exceptions—eg, goods made by prisoners—countries cannot ban imports because they do not like the way they have been produced: if they could, it would give rise to all sorts of protectionist jiggery-pokery. So WTO panels, like GATT panels before them, have regarded production methods as irrelevant: a shrimp is a shrimp, whether or not the net in which it was caught has a turtle-excluder.

Many environmentalists object. If the WTO rules out all trade sanctions that discriminate between more and less conservationist methods of production, then what chance is there of cleaning up? There should be exceptions to allow countries to favour goods made in less polluting ways.

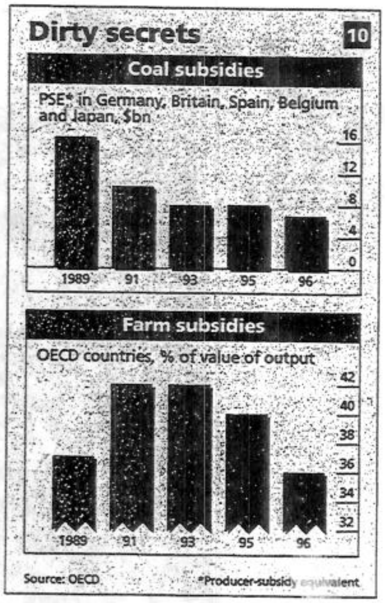
According to Mr Brack, in some circumstances they may be right. Where pollution spans or crosses borders, the threat of trade sanctions can be an efficient way of achieving an environmental goal. The Montreal Protocol is a case in point. Other MEAS might be more effective if their trade clauses were immune from challenge by non-signatories at the WTO.

But this is far from straightforward. Suppose that the signatories to an MEA—to reduce emissions of certain gases, say—were mainly rich countries and the non-signatories were mainly developing ones. The developing countries might be less concerned about reducing pollution, or they might think they were being asked to bear too much of the burden. Would the rich be justified in banning imports from the poor? Clearly any exceptions to Article XX for MEAS would have to be chosen with care.

In any event, where pollution is confined to a single country, the argument for using trade sanctions against it is hard to stand up. Different parts of the world, Mr Brack points out, can tolerate different levels of pollution, depending on climate, the preferences of local people and governments, and existing pollution levels. Imposing the same environmental rules on every country, backed by trade sanctions, would destroy the comparative advantage of many countries, especially developing ones. It would be easy for protectionists to use greenery as an excuse for import restrictions.

Coming out of the shell

The WTO is looking around for ways of accommodating the greens' complaints. For example, it plans to open up its procedures by allowing non-governmental organisations to provide briefs to dispute-settlement panels and to attend hearings. It is already taking more advice from experts: in the shrimp-turtle case, for example, it assembled a clutch of marine biologists. That should help to deflect the environmentalists' charge that the WTO is secretive and lacks know-how.


Dirty secrets 10

Source: OECD *Producer-subsidy equivalent

Changing the WTO's rules, however, is a taller order. In theory, says Mr Jackson, it would be easy to add a paragraph to Article XX to the effect that nothing in the WTO's rules prevents the application of certain MEAS. But he adds: "Politically, I can't see it going through." Such an amendment would have to be approved by three-quarters of all WTO members, and opposition from non-signatories to those MEAS would be robust.

But there are things governments can do that would make trade freer and the planet cleaner at the same time. According to a recent OECD study*, governments continue to support agriculture, energy and transport in ways that damage the environment (see chart 10). Minimum farm prices, for example, encourage farmers to produce too much, using more energy and more chemicals than they should. Coal subsidies encourage power companies to use dirty methods of electricity generation. Such subsidies are already in decline. Speeding their demise would please both free-traders and environmentalists.

---

* "Improving the Environment through Reducing Subsidies". 1998

Brothers up in arms

How not to enforce labour standards

A FORMER World Bank economist recounts a field trip to Morocco. Her guide led her through the medina to watch people working in miserable conditions, buffing up pots for the tourist trade. Next, they ducked into a workshop lit by a single weak lamp where two men were making moulds from dirt. They went on, and down, into more heat and darkness, until they came to the last circle of this little hell: in a near-lightless room, a dull-eyed boy was shovelling coal into a kiln.

> 'How long do you work?' asked the economist.
> 'Ten hours a day, six days a week,' he replied.
> 'How old are you?'
> 'Sixteen.'
> 'And how long have you been working here?'
> 'Nine years.'

The boy was uneducated. His family, back in his home village, needed money. Sadly, his plight is far from the worst in the developing world, and it is anything but new. Indeed, a century or so ago, that economist's European forebears may well have faced the same choice: work or starve.

Trade unions, backed by some politicians, think they can use international trade as a lever to make things better. Use trade sanctions as a punishment for countries that ignore minimum standards of decency, the argument goes, and they will come to their senses. But the WTO disagrees: its rules do not ban imports made by either slave labour or child labour—only those produced by prison labour.

Some WTO members, notably America, have long wanted to link trade liberalisation and labour standards. Others, especially developing countries, smell protection, and not without reason. American unions have long argued that low wages and labour standards in the third world create "unfair" competition for American workers (see next article). But the WTO points out that labour standards are the domain of the International Labour Office (ILO), another organisation based in Geneva.

The ILO, a tripartite body made up of governments, trade unionists and employers from 173 countries, has been going since 1919. It is quite happy to look after its own patch without help from

Labouring 11

ILO conventions on "core" labour standards

No.	Date	Subject	No. of signatories
29	1930	Forced labour	149
105	1957	Forced labour	132
87	1948	Freedom of association	122
98	1948	Right to organise and collective bargaining	138
100	1951	Equal remuneration	137
111	1958	Discrimination	128
138	1973	Minimum working age	61
		Number that have ratified all seven	35

Source: ILO

the WTO. In June it produced a Declaration on Fundamental Principles and Rights at Work which obliges its members "to respect, to promote and to realise" the principles of seven ILO conventions defining core labour standards, even if they have not ratified them (see table 11).

The ILO makes no link between labour standards and trade; indeed its declaration stresses that labour standards should not be used for protectionist purposes. So how will they be enforced? Through shame, hopes the ILO. "In the future," says Michel Hansenne, its director-general, "public opinion will be more important for pushing governments to comply with their commitments." Not good enough, say trade unionists. "The ILO declaration is a great step forward," acknowledges Thea Lee of the American Federation of Labour and Congress of Industrial Organisations (AFL-CIO). "But we'll never see that as a substitute for action in the trade arena. We're not going to give up at the WTO."

The WTO's rules and its dispute-settlement system, trade unionists say, have the bite they are looking for. Bill Jordan, general secretary of the International Confederation of Free Trade Unions (ICFTU), a Brussels-based club of 206 union federations from 141 countries, puts it this way: "We aim to get the world committed to core standards. But we know there's no world government, so we're at a disadvantage. The WTO is the nearest thing we have." The ICFTU wants to see WTO membership made conditional on the observance of core labour standards, certified by the ILO: "If you want to belong to this club," says Mr Jordan, "you have to believe in these standards."

Teeth or gums?

What little information is available suggests that there is no strong link between observance of core labour standards and trade flows, so exporters in developing countries would not suffer much if they adopted the ILO's conventions. By the same token, it seems that workers in rich countries have little to fear from competition with the downtrodden of the developing world.

That is reassuring, but it does not mean that the use of trade measures to enforce core labour standards is a good idea. There is nothing wrong with the intentions behind the standards: civilised people would agree that bonded labour or imprisonment of trade unionists is wicked. But enforcing the standards may not always have the desired effect. If trade unions are recognised, wages in unionised sectors might rise but employment might fall. Displaced workers might be pushed into worse-paid jobs than they had before.

In a paper published last year*, Keith Maskus, an economist at the University of Colorado at Boulder, explains why using trade sanctions to try to improve people's lives may not work. Suppose, he says, that a developing country exports goods made with adult labour, but relies on an input made in an unregulated industry using child workers. Other countries, wishing to discourage the use of child labour, put a tariff on the developing country's goods. Demand for those exports will fall, reducing demand for the adult workers in the industry and cutting their wages. There is also a knock-on decline in demand for its inputs, and therefore the labour of the children making them. So fewer of them work, and their wages fall. Both adult and child employees are worse off.

This could be offset, says Mr Maskus, if the rest of the world recycled its tariff revenue to the target country (or made a contribution from other taxes) in return for a higher minimum working age. The best solution, though, would be not to use trade restrictions at all, but to find other policies that avoid their adverse effects. For instance, children are often put to work because their parents' incomes are too low to make ends meet. So it would help to create more and better-paid jobs for adults—which sounds like a case for freer trade, not more restrictions. Equally, children would be less likely to work if education were more affordable. Subsidising education in developing countries would not only stop some children working; in the long run a more educated workforce would also boost the country's productive potential.

Aware of these pitfalls, some advocates of link-

How best to stop this?

---

* "Should Core Labour Standards be Imposed Through International Trade Policy?" World Bank Development Research Group, Policy Research Working Paper 1817, August 1997

ing trade and labour are shy of calling for tariffs and import bans. "We don't want to go down the sanctions route," says Stephen Pursey, the ICFTU's chief economist. His tentative suggestion for making the link is that if a country persistently failed to meet the ILO's standards, it could be denied the right to bring cases to the WTO's dispute-settlement system. But without the risk of a dispute-settlement case, other countries would feel free to maintain whatever trade barriers they liked against imports from the recalcitrant country, or even to put up some new ones.

The losers would be exporters from the punished country, be they good or bad employers, along with their workers. Better by far to stick with the ILO. If trade unions and their allies in government try to borrow the WTO's teeth, they may find that they bite in the wrong place—if at all.

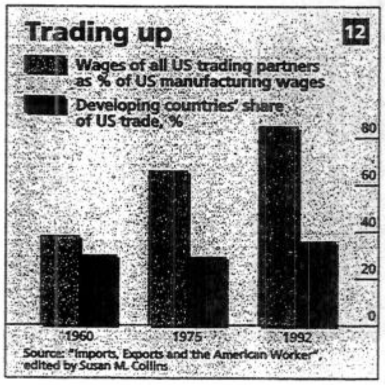
The wages of fear

Are poor countries pinching the rich ones' jobs?

UNTIL the recent jitters, the 1990s seemed to be a golden age for America's economy. Growth was high, inflation and unemployment low. All the same, though, plenty of Americans were wondering why the economy's health was not reflected in their pay packets. Indeed, many Americans have seen little improvement in their living standards in the past quarter-century. After allowing for inflation, average hourly earnings are now lower than they were in 1973—although thanks to increased fringe benefits, average annual earnings have actually risen, according to Susan Collins, an economist at Georgetown University*. At the same time, earnings inequality has increased strikingly. Between 1973 and 1993, the real pay of men 10% of the way up the earnings distribution fell by 21%; for men 10% from the top, it rose by 8%. Women did a little better at the bottom, where earnings went up by 3%, but much better at the top, where they rose by 29%.

In European countries, too, poorer workers are feeling the pinch. Like America, Britain has seen an increase in earnings inequality. In continental Europe, where wages are less flexible, pay differentials have not widened, but unemployment has risen steeply. Some politicians and union leaders, and a few economists, argue that freer trade with developing countries is to blame. This is being said on both sides of the Atlantic, but more loudly in America, especially since the passage of NAFTA. In 1973, only about one-sixth of America's manufacturing imports came from developing countries; now the proportion is one-third. To many people, it seems that American workers' pay is being cut by competition with cheap foreign labour.

But it is not as simple as that. Wages reflect productivity, which implies two things. First, Americans' wage growth has slowed down since 1973 because productivity growth has been lower. Second, the gap between wages in America and wages in developing countries, on average, reflects Americans' higher productivity. So although Americans generally have much fatter pay packets than workers in developing countries, they cost no more to employ per unit of output, and often less. Moreover, says Ms Collins, although America now buys a slightly larger share of its imports from developing countries than it did 25 years ago, average wages in its supplier countries have gone up relative to wages in America, because wages in developing countries have risen much faster than those in rich countries (see chart 12).


Trading up 12

Source: "Imports, Exports and the American Worker", edited by Susan M. Collins

Those same trade unionists and politicians argue that, if some Americans are losing out from freer trade, America should put up barriers against imports from poorer countries. Trade with Mexico and China is a favourite target. Yet import barriers would make Americans as a whole worse off: American companies would be under less pressure to make goods as cheaply as they can, and resources would be wasted on making things that could be bought more cheaply from abroad. Everything would get more expensive—and poor Americans as well as rich ones would have to stump up.

A price to pay

Nevertheless, there are good reasons why freer trade might hurt some of America's poorer workers. The whole point of free trade is that it makes countries specialise in industries where they have a comparative advantage; so when trade barriers are removed, resources are shifted into those industries, and away from less productive uses. In America, although this change undoubtedly benefits the economy as a whole, it may work to the disadvantage of unskilled, and therefore low-paid, workers.

As any economist will tell you, this phenomenon can be explained by the Stolper-Samuelson theorem, which says that a fall in the price of a good—say, as a result of a tariff cut—reduces the pay of the type of labour most heavily used in making it. So a fall in American tariffs on imports from Mexico of goods that use low-skilled labour relatively intensively will cut the wages of low-skilled Americans.

Do the facts fit the theory? Economists have concentrated on explaining the increase in the "skill premium", measured in most studies by the ratio of the pay of workers with college degrees to that of high-school graduates. In the 1970s this was around 1.3-1.4, but since 1980 it has risen fairly steadily, to about 1.7 now. Some studies have concluded that trade explains a lot of the skill premium, but most find the effect of trade is rather modest, accounting for perhaps one-fifth of the rise in the premium. This is partly because the evidence on prices does not quite fit the story: the relative prices of goods in low-skill industries did not fall in the 1980s, when

---

* This article draws on a book edited by Susan M. Collins, "Imports, Exports and the American Worker". The Brookings Institution Press, Washington, DC, 1998

UAW
ON
STRIKE
STRIKE
UAW

At the losing end

the skill premium rose sharply, but they probably did in the 1970s, when it was constant.

Other factors seem to have been more important. Richard Freeman, of Harvard University, reckons that immigration has probably had a far bigger effect on the market for low-skilled labour than trade. Workers in many service industries, from barbers to taxi drivers, are fairly well insulated from international trade, but will meet hot competition from immigrants.

Most important of all seem to have been advances in technology, which have favoured the better-educated at the expense of the rest. Why only "seem"? Because the effects of technical change are hard to measure directly. Most models explain as much as they can of the change in the skill premium in terms of price changes, trade flows, immigration and whatnot, leaving a large chunk unexplained. That residual is then attributed mainly to changes in technology.

The effects of changes in technology and changes in trade are hard to unravel. Suppose that a foreign company discovers some new cost-cutting technique for making widgets, which brings down the price of widget imports into America. If the wages paid by its American competitors fall as a result, are the workers victims of trade or technology?

If trade has a smallish effect on the skill premium, it has an even smaller effect on wage inequality as a whole, says Ms Collins. Only about one-third of the increase in inequality can be put down to differences between workers—whether in education, race or sex. So at most only 7% of the increase in wage inequality can be blamed on trade.

Never mind the economics, count the votes

None of this, however, has stopped freer trade from becoming a political problem. Is it always possible to find anecdotes about jobs lost to foreign competition, or of workers swallowing pay cuts to stop their bosses from moving their factory to Mexico. Moreover, the benefits of trade tend to be more spread out than the losses. The local effects of a factory closure are only too obvious; the across-the-board effects of greater consumer spending power generated by cheaper imports, or of higher productivity brought about by increased competition, are far less evident.

One way of getting round these political problems is to compensate the losers from freer trade. In practice, that means those whose jobs disappear rather than those who merely suffer pay cuts. Since freer trade should benefit the economy as a whole, some of that benefit could be channelled to the losers through the tax and welfare systems.

This idea has, in fact, been put into practice in America since the Trade Expansion Act of 1962. Workers displaced by foreign trade can apply for Trade Adjustment Assistance (TAA). This extends unemployment insurance payments for up to a year beyond the usual six-month period, provided that claimants take part in a training programme. Since 1974, about 2m workers have been certified as eligible for TAA. The number taking it up has been much lower, because many of the workers have found other jobs.

In a recent book, "Globaphobia"*, four American economists—Gary Burtless, Robert Lawrence, Robert Litan and Robert Shapiro—say that TAA should be reformed. The payments go on for too long, removing the incentive to get a job quickly. Workers who find a new job at lower pay should also be compensated, to give them an incentive to work for less. And the training does little to fit workers for new jobs, so it should be scrapped.

All this would be an improvement on the current system, but two problems remain. First, how can the losers from trade be identified? Just think of the difficulties of unravelling the effects of technology and tariff cuts; then remember that trade is about a lot more than border measures. What about deregulation to allow foreign ownership of American airlines, shipping or television stations; or recognition of European drug-testing regimes, or of foreign medical qualifications?

Second, what is so special about trade? It is just one of many forces that send resources whizzing around the economy. It seems odd to pay a worker compensation because his firm has succumbed to competition from imports, but give him nothing if it has been put out of business by a company up the road, or failed to adapt to changes in technology or in customers' tastes. Paying out in this way seems to support protectionists' contention that trade is a blight on employment like no other.

The four authors are aware of both these drawbacks. They suggest getting round the first problem by limiting compensation to victims of specific trade agreements, and by allowing Congress to specify the industries in which compensation will be paid. Yet as they note, the extremely complex Uruguay round triggered no special adjustment measures, even though it will have a far greater effect on the American economy than NAFTA, which did prompt such measures. The second point—that trade is just one of many economic forces, and should not be singled out—is unanswerable.

---

* Published by the Brookings Institution Press, the Progressive Policy Institute and the Twentieth Century Fund

But perhaps such carping misses the point. Adjustment programmes are political devices, not economic ones. And for free-traders, the most bewildering fact about the American economy is that, despite all the stability and success in recent years, Americans have remained nervous about freer trade. So have plenty of their elected representatives. Why?

Slow road to fast-track

Does it matter?

WHEN Bill Clinton addressed the WTO in Geneva last May, he issued an invitation to trade ministers: next year, come to America. Some governments hope that this meeting, due to start on November 30th 1999, will be the overture to a new round of global trade talks. The European Union's Sir Leon Brittan has long been calling for a "millennium round". Yet the Clinton administration has been shy of endorsing a new round. Why?

America's reticence is partly tactical. In the past, the Americans argue, rounds have dragged on and on—the Uruguay round lasted an epic seven-and-a-half years—because nothing is decided until everything is decided. So before starting a new round, American officials want to set a timetable for completing talks on different subjects and implementing their results.

However, there is another reason for America's shyness. Mr Clinton does not have "fast-track" negotiating authority from Congress, which has the constitutional right to "regulate commerce with foreign nations and between the several states". Fast-track authority allows a president to negotiate trade agreements and ask Congress to pass the necessary legislation within 90 working days and without amendment. Without fast-track, trade agreements can be put before Congress as ordinary bills, which can be amended, or they can be presented as treaties. But treaties require approval by two-thirds of the Senate—where they can also be amended or held up indefinitely by the foreign-affairs committee. The beauty of fast-track authority is that it prevents Congress from rewriting trade deals once the president has signed them, and therefore boosts his credibility in trade negotiations.

EMERSON'S
BROMO-SELTZER
INSTANTLY CURES
HEADACHE

All Headaches Instantly Cured
or Money Refunded.

LEGAL GUARANTEE.

6D. EMERSON'S BROMO-SELTZER, the most successful American Remedy, is an effervescent Powder, taken in water. If three doses do not Cure any Headache, no matter how caused, send the Bottle to us, saying where obtained, AND WE WILL AT ONCE REFUND THE PRICE. TRIAL BOTTLE, post free, 6d. Larger Sizes 1s. and 2s. Sold by many Chemists or obtained to order by almost all.

EMERSON DRUG CO., LTD.,
46, HOLBORN VIADUCT, LONDON, E.C.

Insist on Full Name—

EMERSON'S BROMO-SELTZER

A persistent headache

For most of the past quarter-century, presidents have had few problems winning fast-track authority. From free-traders' point of view, that has been immensely useful. It has helped to push through five big trade agreements: the Tokyo round of GATT, free-trade agreements with Israel and Canada, NAFTA, and the Uruguay round, the last two during Mr Clinton's first term in the White House. But since 1994 Mr Clinton has been without fast-track authority. Last autumn he was on the point of formally asking for it, but when it became clear that he would not get it no vote was taken.

Curmudgeonly Congress

Given the rude health of the American economy over the past few years, the administration's failure to win fast-track authority seems remarkable. It is more surprising still in view of Mr Clinton's success, in his first year in office, in squeezing NAFTA through Congress by forging a coalition between Democrats and Republicans.

So why has he failed? In essence, because Congress is deeply divided over what a fast-track bill should say about labour standards and the environment. Most Republicans think that trade should not be linked to these issues at all. Most Democrats think it should.

Since the Democrats are in a minority in both houses, there is no chance in this Congress of passing a bill encapsulating their ideas on a trade mandate. Hence Mr Clinton's attempt last year to win fast-track authority on the Republicans' terms. Such a bill would have excluded clauses on labour and the environment from fast-track treatment. However, despite the Republican majority the bill could not pass without some Democrat votes because not all Republicans support free trade.

Mr Clinton failed to persuade members of his own party to support a "Republican" bill. "The Democrats think Clinton left them a long time ago," says Mac Destler, a specialist in American trade politics at the University of Maryland. Paradoxically, this lack of trust has its roots in the president's success in building a pro-NAFTA coalition in 1993. At the time most Democrats were opposed to the agreement, which Mr Clinton inherited from George Bush. They worried that American workers would be hurt by freer trade with Mexico because Mexico's looser labour and environmental standards would give it an unfair advantage. Mr Clinton put their minds at rest by adding side accords on labour and the environment. And to win over Hispanic Democrats, he came up with a North American Development Bank (NADBank) to finance environmental projects on the Mexican border and to help workers who lost jobs to new competition from Canada and Mexico. That brought enough Democrats on board.

Four years on, however, those Democrats felt short-changed. The labour and environment side accords never amounted to much, merely obliging each NAFTA country to respect its own laws. The NADBank has, until recently, done little. And among the losers from NAFTA, Hispanics stood out: Raul Hinojosa-Ojeda, of the University of California at Los Angeles, reckons that many Mexican immigrants who had headed north in search of work saw jobs move in the opposite direction. Thus last year, when Mr Clinton was looking for fast-track votes, congressional Democrats insisted that they could not support a bill denying fast-track treatment to labour and environment clauses. It was weaker in that respect, they said, than the authority given to Mr Bush.

In this Congress, fast-track is not quite dead. Newt Gingrich, the Speaker of the House of Representatives, is pushing for a vote on last year's aborted bill. This would simultaneously endear him to the farm lobby, which is keen on fast-track, and make mischief for the president, who would rather wait until 1999. Mr Clinton has no wish to advertise his own party's divisions on trade so soon before congressional elections, and for now would rather use such political capital as he still has on the equally fraught issue of financing the International Monetary Fund. And without the administration's backing, a vote on fast-track is unlikely to be carried.

The president's chances of winning fast-track in the next Congress look better, if still not good. If the balance between the parties after the mid-term elections later this autumn stays much the same as it is now, he may try again to find some wording that will satisfy enough Republicans and enough Democrats. This looks his best bet: Hispanic Democrats, in particular, have sounded more emollient in recent months, now that the NADBank has at last started disbursing money.

Next year, some time, never?

How much do these congressional shenanigans matter to getting trade agreements into place? Not much, says the administration. Charlene Barshefsky, the United States trade representative, reels off a long list of all the deals that have been struck without fast-track, including the WTO's agreements on telecommunications, financial services and information technology. The Europeans also say they are not fussed: "You need fast-track to complete the round, not to start it," explains Sir Leon. Ms Barshefsky and Sir Leon seem to have precedent on their side. The Tokyo round began in 1973, but Congress did not give the administration fast-track authority until 15 months later. And when President Reagan was granted fast-track in 1988, the Uruguay round had already been going for two years.

The difference, says Craig VanGrasstek, a Washington trade-politics expert, is that in 1973 and 1986 there was not much doubt that negotiating authority would be granted, nor about the likely terms. Now no one really knows. In the mid-1980s, there was a congressional coalition in favour of negotiating trade agreements on the then-new issues of services trade and intellectual property, where American exporters had plenty to gain and domestic firms had little to fear from foreign competition. But now labour standards and the environment are so divisive that building a new coalition may prove impossible.

Barshefsky: fast-track is not a worry

In the absence of fast-track authority, the president has limited scope to strike new deals without asking Congress's approval. In 14 sectors where America made "zero-for-zero" tariff offers in the Uruguay round, the president is still free to cut tariffs in WTO negotiations. These include farm equipment, electronics, drugs, steel and toys. This freedom is being used in talks on tariff reductions by APEC countries, which are meant to apply to imports from non-APEC countries as well. But in most of the sectors under discussion, including energy and fish, the president does not have authority to cut tariffs.

In essence, he is free to make deals only if they require no changes in American law (which would require Congress's permission). This proved a serious constraint to America in the WTO talks on financial services and telecommunications, says Mr VanGrasstek. With more freedom to offer concessions of their own, he suggests, the Americans might have been able to prise more out of their negotiating partners. America's limited room for manoeuvre will also prove unhelpful in future talks. Chile has already made it clear that it is not prepared to discuss joining NAFTA until fast-track is in place (but has meanwhile struck deals with Canada and Mexico). Small beer, perhaps; but nevertheless a blow to Mr Clinton's planned Free-Trade Area of the Americas, and a sign that WTO talks without fast-track could be sticky.

portant decision they will face will probably be whether to launch a new round. Their answer should be an unreserved yes. There are at least four reasons why.

The first is that trade negotiations seem to work best when a range of issues can be bundled together. Single-issue negotiations on services have not been an unqualified success. Negotiations on maritime services were sunk by the Americans before they left port. Rich-country indifference also dashed any hopes of an agreement on the "movement of natural persons", trade-speak for the temporary employment of foreign workers. A round that ties together a complex package of issues offers a better chance of striking deals: concessions in one area can be traded for advances in another.

Second, a long agenda is building up. In addition to services and agriculture, there are manufacturing tariffs, the treatment of foreign investors and competition policy. Sooner or later, too, WTO members will have to find some way of resolving the conflict between the organisation's rules and those of multilateral environmental agreements.

A round covering all this would go some way to completing the huge shift in trade rules begun in the Uruguay round. These rules now recognise that free trade amounts to far more than the removal of barriers to trade in goods. But in services and investment, foreign suppliers are often still a long way from getting "national treatment". Some agreement on antitrust law would help to stop private trade barriers replacing official ones.

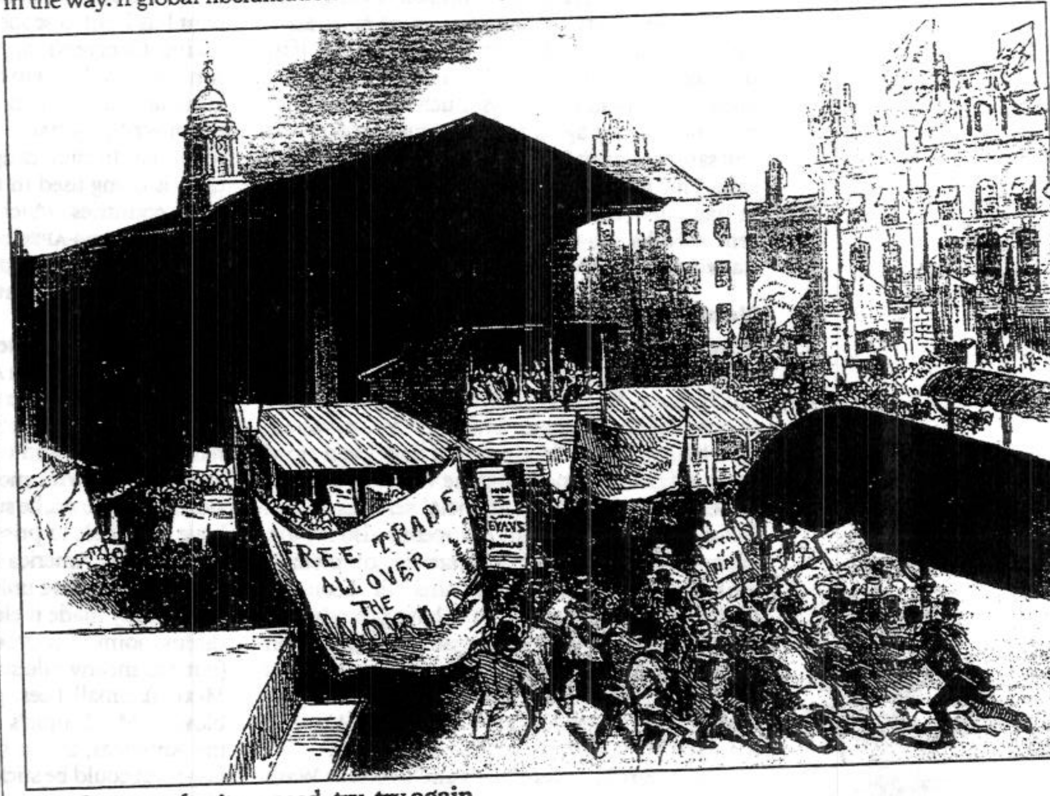
The third reason is that if the WTO does not come to grips with the new issues, regional trade organisations will continue to go their own way. In some cases that may be fine, but in others it may get in the way. If global liberalisation were continuing apace, there would be less need to worry about the way regional agreements distort trade.

Why now is a good time

The fourth and most important reason for a new trade round is the parlous state of the world economy. This argument may not be immediately obvious. The deep recession in many Asian economies and the turmoil in world financial markets may well spread to the trading system. Already, Americans and Europeans are struggling to sell their wares in depressed Asia, and are losing markets to Asian goods made far more competitive by huge devaluations. Some westerners may, therefore, be tempted to call for curbs on trade. And if financial markets crash, or the world economy moves into recession, free markets may seem less of a good idea than they did a couple of years ago.

Yet these are arguments not for closing the shutters, but for opening them more widely. In the 1930s, protection turned recession into rout. If the same remedy were adopted, history could repeat itself in the 1990s. Freer trade, on the other hand, offers the prospect of earlier recovery.

Helpfully, WTO commitments bind governments, up to a point, to keeping markets open, come rain or come shine. Moreover, this would not be the first time that a trade round started in turbulent times. Fred Bergsten, of the Institute for International Economics, points out that the Tokyo round was launched in 1973, in the middle of the first oil crisis; and the preliminaries for the Uruguay round came in the wake of the Latin American debt crisis. Launching a new round in 1999 would be a signal that governments are still committed to open markets—and understand that the prosperity of their peoples depends on them. ■



If at first you don't succeed, try, try again

Reprints
Reprints of this survey are available at US$3.75 each, plus 10% postage. Add tax in CA, DC, IL, NY, VA; GST in Canada.
For classroom use or quantities over 25, please telephone for discount information.
Please send your order with payment by cheque or money order to The Economist Newspaper Group, Inc.,
Reprints Department,
111 W 57th St,
New York, NY 10019.
Telephone (212) 541-5730
(American Express accepted)

24

WHARTON REPROGRAPHICS

FOREIGN PROFITS AND THE POLITICAL ECONOMY OF QUOTAS

Paul E. Godek

> Tariffs are the weak, quantitative restrictions, the strong form of government interference with international trade.
>
> —Jan Tumlir (1985)

Introduction

Trade restrictions subsidize domestic producers. Foreign producers, along with consumers and allocative efficiency, are traditionally viewed as losers in the process of redistribution through trade restrictions. The increasing use of quantity restrictions, however, has induced students of trade policy to re-examine part of the traditional approach. New protectionist measures often take the form of discriminatory quotas—quotas allocated to individual foreign producers.[1] Such quotas can make foreign producers beneficiaries of protectionism by giving them a property right similar in value to the forgone revenue of the equivalent tariff. Here, I examine the simple relationship between quotas and the profits of foreign producers in order to demonstrate the surprising range over which a quota can raise foreign

---

Cato Journal, Vol. 11, No. 1 (Spring/Summer 1991). Copyright © Cato Institute. All rights reserved.

The author is a Senior Economist at Economists Incorporated in Washington, D.C. He thanks Michael Baumann, Ken Heyer, Kent Mikkelsen, Larry Sjaastad, and David Smith who provided helpful comments, and Diane Letourneau who provided expert research assistance. The usual caveat applies.

[1] Nontariff barriers of some sort cover between 30 percent and 50 percent of world trade by some estimates. See Tumlir (1985) for a summary of studies examining the growth of nontariff trade restrictions, as well as for an incisive study of the political and legal history of protectionism. The U.S. experience is well documented by Bergsten et al. (1987). They estimate that 18.3 percent of U.S. imports were covered by quantitative restraints in 1986, up from 5.1 percent in 1980. When quotas are instituted and allocated by the exporting country, they are often referred to as voluntary export restraints.

producers' profits above free-trade levels.[2] The results reveal one of the political functions served by quotas and may help explain their increasing use.

Quotas and Profits

In general, quotas are assigned, not auctioned off. They may be granted directly to foreign producers or to individual countries, which then allocate them to producers. In either case, quotas are usually assigned to foreign producers on the basis of their historic market share. Hence, quotas transfer to the foreign producers the forgone revenue of an equivalent tariff. Such quotas are, of course, more profitable than the equivalent tariff to foreign producers. It is revealing to examine the range over which quotas are more profitable to foreign producers than free trade.

As long as the slope of the residual demand curve facing foreign producers is less than infinite—that is, as long as the elasticity of supply of domestic producers is less than infinite—foreign producers, as a group, face a downward-sloping demand curve, and their profit-maximizing output is less than the competitive (free-trade) output. That is trivial. But how much can imports be reduced before foreign producers' profits fall below the free-trade level?

A simple example demonstrates the magnitudes that might be involved. The supply functions of foreign and domestic producers are constructed so that import supply is slightly more elastic than domestic supply, and under free trade, imports have 40 percent of the market:

(1) Market Demand: $d = 30 - p$

(2) Domestic Supply: $s = -3 + 1.5p$

(3) Import Supply: $i = -12 + 2p$

(4) Total Supply: $t = i + s.$

Where d is quantity demanded, p is price, s is domestic supply, i is quantity imported, and t is total supply. Under free trade, the equilibrium price is 10, quantity demanded is 20, domestic supply is 12, and import supply is 8. Figure 1 displays the model.

---

[2] Bergsten et al. (1987) present a comprehensive study of the effects of all U.S. quantitative restraints on foreign producers' profits. They estimate that quotas currently in place transfer over $9 billion to those holding the rights to import. The effect of quotas on foreign producers' profits also receives explicit attention in Bhagwati (1989), Deardorff (1987), Kaempfer et al. (1987), Hufbauer et al. (1986), Godek (1985), Clements and Sjaastad (1984), Ono (1984), and Jones (1984). Denzau (1986), Hickok (1985), Tarr and Morkre (1984), and Crandall (1984) are industry specific studies that examine quotas and foreign producers' profits.

Now replace import supply with a quota. With linear supply and demand, foreign producers' profits are a quadratic function of the quota level. In this example, the profit (π), quota (q) function is

$$\pi = -.65q^2 + 7.2q. \tag{5}$$

Figure 2 displays foreign producers' profits as a function of the quota level. Foreign profits are computed using the import supply function as the marginal cost function. The quota level is displayed as the ratio of allowed imports to the free-trade level. That ratio ranges from one (free trade) to zero (a complete prohibition of imports).

FIGURE 1

FREE-TRADE MODEL

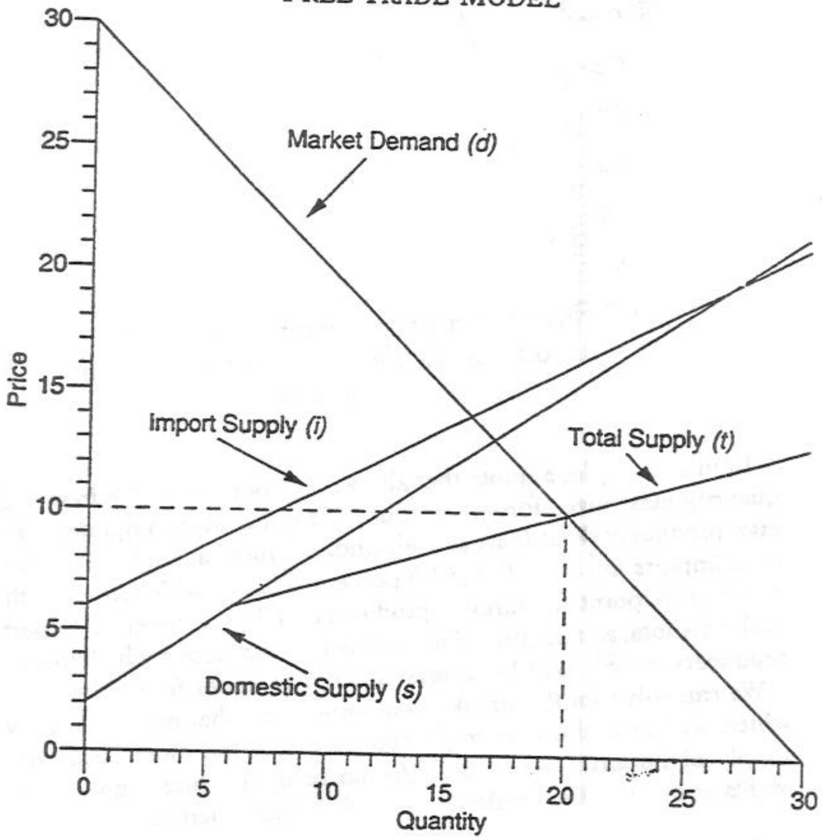
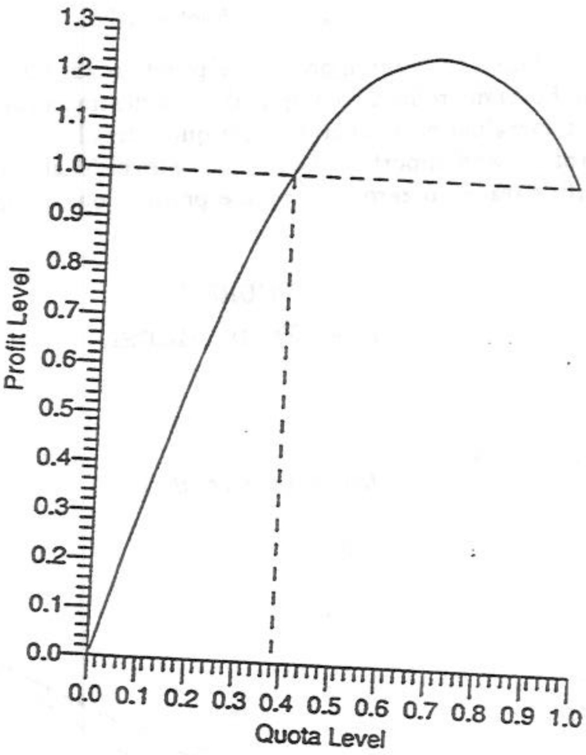


FIGURE 2

FOREIGN PRODUCERS' PROFIT LEVEL AND QUOTA LEVEL



In this example, a quota that allows 69.2 percent of the free-trade quantity maximizes foreign producers' profits. More important, foreign producers do at least as well under a quota as under free trade until imports fall more than 61.5 percent. That is, 38.5 percent is the break-even point for foreign producers. It is the level of imports under a quota, as a fraction of free-trade imports, above which foreign producers' profits will be at least as high as under free trade.

We can solve for the break-even point somewhat more generally, which will also allow us to demonstrate that the more elastic the supply of imports, the greater the range over which quotas raise profits above the free-trade level. With linear functions,

(6) $$d = A + Bp$$

(7) $$s = C + Dp$$

(8) $$i = E + Fp.$$

Profits as a function of the quota level are

$$\pi = -\{[1/(D-B)] + [1/(2F)]\}\,q^2 + [(A-C)/(D-B) + (E/F)]\,q. \tag{9}$$

The range, over which profits are above the free-trade level, is defined by the smaller root of the quadratic equation formed by subtracting the free-trade level of profits from the right-hand side of the above equation.

Let p_f represent the free-trade price and q_f the free-trade quantity of imports. The quadratic can then be written:

$$-\{[1/(D-B)] + [1/(2F)]\}\,q^2 + [(A-C)/(D-B) + (E/F)]\,q - \{(p_f q_f) - [(q_f)^2/(2F)] + [q_f(E/F)]\} = 0. \tag{10}$$

The constant term is the free-trade profit level. We can use the free-trade equilibrium condition $p_f = (A - C - q_f)/(D - B)$ and the identity $E = q_f - Fp_f$ to simplify the equation to

$$-\{[1/(D-B)] + [1/(2F)]\}\,q^2 + \{(q_f/F) + [q_f/(D-B)]\}\,q - [(q_f)^2/(2F)] = 0. \tag{11}$$

The smaller root of that quadratic, as a fraction of q_f, is

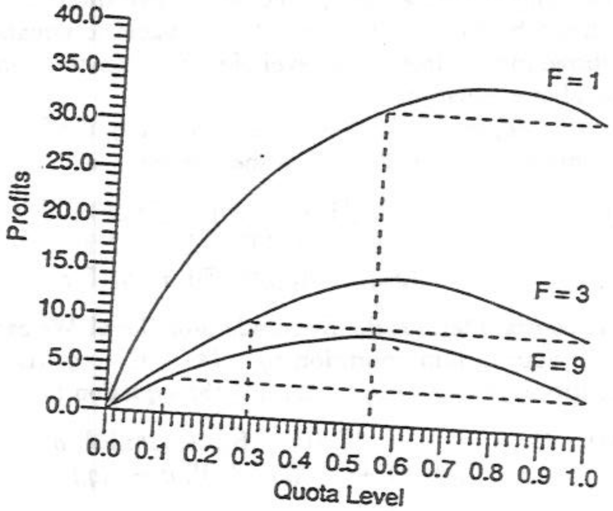
$$(D-B)/(2F + D - B). \tag{12}$$

That value is the break-even point for foreign producers. Once again, the break-even point is the level of imports under a quota, as a percentage of free-trade imports, above which profits will be greater than under free trade. Note that that value is a simple function of the slopes and not the intercept terms. For example, regardless of free-trade import shares, if the slope terms are all approximately equal in absolute value, foreign producers' profits increase under a quota unless imports are reduced by more than 50 percent.

Note also how the slope of the import supply function (F) affects the break-even point. The break-even point decreases with F, approaches zero as F approaches infinity, and approaches one as F approaches zero. Thus, for any given domestic supply and demand functions, the more elastic the supply of imports, the larger is the range over which quotas raise foreign profits over the free-trade level. Figure 3 displays the relationship between quota levels and foreign producers' profits for three different values of the import supply response. The values of F used in the figure are 9, 3, and 1. (For clarity, the value of E is adjusted to hold the free-trade import share at 40 percent. Note that the horizontal location of the break-even

FIGURE 3

FOREIGN PRODUCERS' PROFITS AND QUOTA LEVEL AT VARYING IMPORT ELASTICITIES



point is not affected by whether import shares are held constant in this way. The domestic demand and supply functions are those used in Figure 2.) Only as the import supply response approaches zero does the profit level decline monotonically with the quota level. As the supply response approaches infinity, a foreign producer will prefer any quota level (above zero) to free trade.

While the exact results here depend on linearity, the basic finding is quite general. In many markets, there is likely to be a large range over which quotas can be enforced while increasing the profits of established foreign producers. Given that the market shares of foreign producers are usually not dramatically reduced by quotas, the large rent transfers reported by Bergsten et al. (1987) should not be surprising. With quotas, the political sector can join foreign and domestic producers in an alliance of protectionism, even when imports are significantly reduced.

Political Equilibrium

How do international trade quotas fit into the economic theory of regulation? Following the approach of Stigler (1971) and Peltzman (1976), we explain government policy as a political-support-maximizing tradeoff among competing groups with different lobbying efficiencies—that is, among groups with different levels of influence.

From that perspective, an import restriction is viewed as a politically efficient method of redistributing rents or market surplus from consumers to producers. The political equilibrium is reached at the quota level that maximizes political support.[3]

The maximization problem representing the political process is easy to describe. The function that relates foreign and domestic profits and consumer surplus, as determined by the quota level, is referred to as the surplus transfer function. A political preference function, which defines political utility or support as a function of domestic profits, foreign profits, and consumer surplus, is maximized subject to the constraint of the surplus transfer function. The political preference function would be pictured as an indifference set (a set of iso-political-utility surfaces in consumer surplus, foreign profit, and domestic profit space) maximized subject to the surplus transfer function (a curved line in that space).

Figure 4 illustrates the equilibrium in two dimensions, ignoring the level of consumer surplus. The figure shows the surplus transfer function for foreign and domestic profits and the maximum attainable level of the political preference function. (In the figure, profits are measured relative to their free-trade levels, so that both values equal one at the free-trade point.)

An important question is: What particular characteristics of the domestic and international markets explain the shape of the political preference function? That is, what makes some industries more powerful than others? For example, why is copper traded relatively freely, while steel imports are constrained by quotas? And why is there a quota on Japanese automobiles, while Japanese motorcycle and light-truck producers face high tariffs? Those questions have received a good deal of attention from economists, although distinguishing between tariffs and quotas is a relatively new line of inquiry.[4]

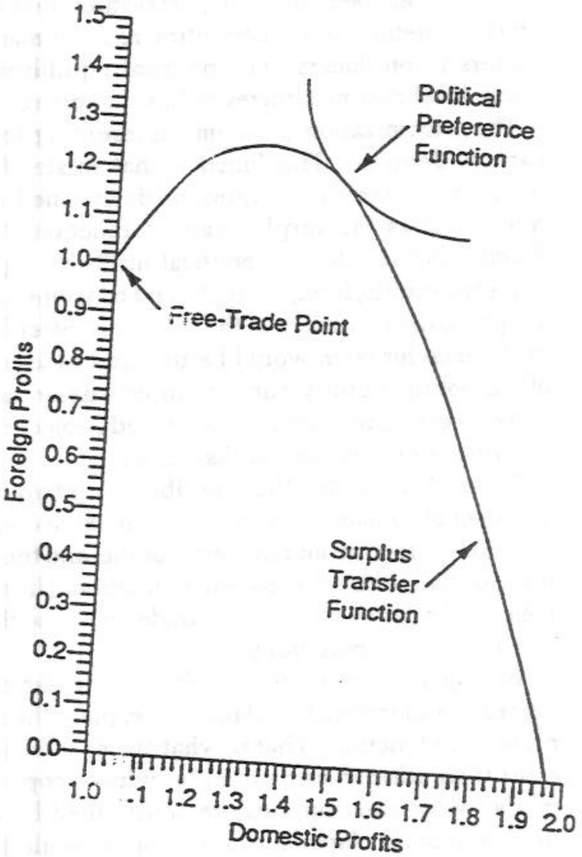
---

[3] Hillman and Ursprung (1988) present a rigorous model of the political economy of alternative forms of trade restriction. Their paper also contains a useful survey of the literature.

[4] For empirical studies of the relation between domestic industry characteristics and types of trade restriction, see Godek (1985), Ray and Marvel (1984), Marvel and Ray (1983), Ray (1981), and the papers they cite. Magee et al. (1989) is a useful survey. For example, as summarized in Godek (1985, p. 700): "Quotas do appear to be a distinct regulatory entity. They are more likely to exist and be more restrictive in larger or more geographically concentrated industries. Tariffs, on the other hand, decrease with the number of firms and increase with the share of the market dealing in final goods. The incentive to generate either type of trade restriction is strongly and inversely related to the average wage." The influence of foreign producers has not yet been controlled for in any systematic way.

FIGURE 4

POLITICAL EQUILIBRIUM



The simple presumption here is that, like the domestic producers, foreign producers may be able to directly influence the political process. Foreign producers may also be able to lobby their governments to intercede on their behalf. Or, as will be discussed, they may be able to bring legal action under international trade agreements to oppose tariff increases. Thus, foreign producers will have varying degrees of influence over the type and level of trade restrictions they face. And as we have seen, if foreign producers are to face a quota, they may not be ardent supporters of free trade.

If we refer to Figure 4 again, corner solutions are possible although rare in practice. If consumers have no political influence (have no weight in the political preference function), imports still may not be totally prohibited. Only when consumers and foreign producers both lack political influence will no imports be allowed. If foreign producers are very influential, relative even to domestic producers, quota levels may approach the foreign producers' profit-maximizing level. If domestic and foreign producers have no political influence, trade will be unencumbered.

The political sector can also choose the surplus transfer function generated by a tariff. In that case, foreign profits and consumer surplus both decline monotonically with imports. A tariff may be used when foreign producers have no political influence or if quotas are particularly difficult to administer in a given market. One of the political benefits of a tariff is, of course, the revenue collected.

However, the imposition of tariffs is more effectively constrained by international trade agreements than are quotas. The case has been made that quotas are used specifically to circumvent those agreements. Tumlir (1985) has argued that discriminatory quotas are used to benefit foreign producers because those producers could successfully challenge increased trade restrictions under the General Agreement on Tariffs and Trade.

The prohibition of high tariffs by international trade agreements may explain the increased use of quotas in recent years. The trade agreements lowered tariffs below their political equilibrium levels. That is, many industries that found themselves with less tariff protection still had the influence necessary to generate a replacement, the discriminatory quota. And the replacement allowed the industries to add another group, the foreign producers, to the lobbying coalition. From that perspective, the increased use of quotas is the ironic result of trade agreements that effectively prohibited high tariffs.[5]

Also, it has been observed that quotas are often used to supplement tariffs. That is, industries that already enjoy a high degree of tariff protection are more likely to be further aided by nontariff barriers.[6] That observation is consistent with the explanation that quotas are used to circumvent international trade agreements. After such agreements, industries with lowered but still higher-than-average tariffs receive further protection with quotas. Similarly, it may be that a tariff is used initially to grant protection. If further protection is required to reach political equilibrium, a quota may then be used, both to appease the foreign producers and to avoid a tariff in excess of agreed-upon levels. In either case, it should be observed that the quota makes the tariff redundant, not vice versa.

---

[5] That analysis implies that the separate industry lobbying groups were unable to form a single coalition in order to stop the trade agreements.

[6] See the papers cited in footnote 4, in particular Godek (1985, pp. 693–701).

Marvel, Howard P., and Ray, Edward J. "The Kennedy Round: Evidence on the Regulation of International Trade in the United States." *American Economic Review* 73 (March 1983): 190–97.

McCulloch, Rachel. "Why Do Governments Prefer Nontariff Barriers? A Comment on Deardorff." *Carnegie-Rochester Conference Series on Public Policy* 26 (Spring 1987): 217–21.

Ono, Yoshiyasu. "Profitability of Export Restraint." *Journal of International Economics* 16 (1984): 335–43.

Peltzman, Sam. "Towards a More General Theory of Regulation." *Journal of Law and Economics* 19 (August 1976): 211–40.

Ray, Edward J. "The Determinants of Tariff and Nontariff Trade Restrictions in the United States." *Journal of Political Economy* 89 (February 1981): 105–21.

Ray, Edward J., and Marvel, Howard P. "The Pattern of Protection in the Industrialized World." *Review of Economics and Statistics* 66 (August 1984): 452–58.

Stigler, George J. "The Economic Theory of Regulation." *Bell Journal* 2 (Spring 1971): 3–21.

Tarr, David G., and Morkre, Morris E. *Aggregate Costs to the U.S. of Tariffs and Quotas on Imports: General Tariff Cuts and Removal of Quotas on Automobile, Steel, Sugar, and Textiles.* Washington, D.C.: Federal Trade Commission Staff Report, December 1984.

Tumlir, Jan. *Protectionism: Trade Policy in Democratic Societies.* Washington, D.C.: American Enterprise Institute, 1985.

LIPPINCOTT LIBRARY
DOCUMENT DELIVERY

WHARTON REPROGRAPHICS

Southern Economic Journal 1997, 64(2), 402–424

Invited Article

Dumping on Free Trade: The U.S. Import Trade Laws

Joseph E. Stiglitz*

The passage of the Uruguay Round implementing legislation represents a natural opportunity to review the policy goals of the U.S. import trade laws, to assess how well current laws achieve those objectives, and to explore possible reforms. I argue that there is a variety of policy concerns justifying a circumscribed set of import trade statuses. The relevant U.S. laws, however, have largely become divorced from such national welfare considerations and are now too often a mechanism for furtive protectionism. The Uruguay Round effected some (marginal) improvements but left the fundamental structure of the laws unchanged. I discuss possible reforms in the final section of the paper.

1. Introduction

As import tariffs have fallen across the world, nontariff barriers have become more important impediments to trade. The use of unfair trade laws, in particular, has increased dramatically over the past two decades—undoubtedly much more dramatically than any increase in unfair trade itself. In the United States, for example, 233 antidumping cases were investigated between 1968 and 1978, whereas almost twice that number (451 cases) were investigated between 1980 and 1989.[1] Unfair trade laws are also being used more frequently abroad: Figure 1 shows the number of antidumping cases filed by non-U.S. GATT participants since 1980 (note the apparent counter-cyclical pattern of the series).[2]

In December 1994, President Clinton signed the Uruguay Round implementing legisla-

---

* Senior Vice President and Chief Economist, World Bank, 1818 H Street, NW, Washington, DC 20433, USA; e-mail JSTIGLITZ@WORLDBANK.ORG.

I thank John Anderson, Bob Cumby, Gary Horlick, Michael Meyer, Louise Keely, Caroline Levine, and Michael Orszag for helpful comments and/or data (subject to the usual disclaimer). I am especially grateful for the contributions of Peter Orszag, who developed many of these ideas with me during our time together at the Council of Economic Advisers and subsequently while he was completing his dissertation at the London School of Economics. The opinions expressed here do not necessarily represent those of the Council of Economic Advisers or any other agency of the U.S. government.

[1] See Dale (1980, Appendix 3, p. 204) and Boltuck and Litan (1991, Table 1-1, p. 2). The discrepancy obtains despite the difference in time intervals (the first figure covers an 11-year period, whereas the second figure covers a 10-year period). Adjusting for the increased volume of merchandise imports during the later period attenuates, but does not eliminate, the increase. In the earlier period, 11.9 cases were investigated for every $100 billion (1987 dollars) of merchandise imports. In the later period, 13.2 cases were investigated for every $100 billion of merchandise imports. Thus, if unfair imports are assumed to be a constant proportion of total merchandise imports, there was still a significant increase in the use of antidumping procedures, even after accounting for the increased incidence of unfair trade.

[2] Figure 1 depicts the total number of antidumping cases reported to the GATT Committee on Antidumping Practices minus the number reported by the U.S. Only signatories to the antidumping code within the GATT are required to file such reports, and many countries (including many GATT Contracting Parties) are not signatories. In addition, some signatories (e.g., the European Union) do not report some cases filed against nonsignatories. The figures presented thus represent a lower bound on the total number of antidumping cases filed outside the U.S. Sources: For 1980 through 1988, U.S. General Accounting Office (1990, p. 16). For 1989 through 1993, compilations based on GATT reports.

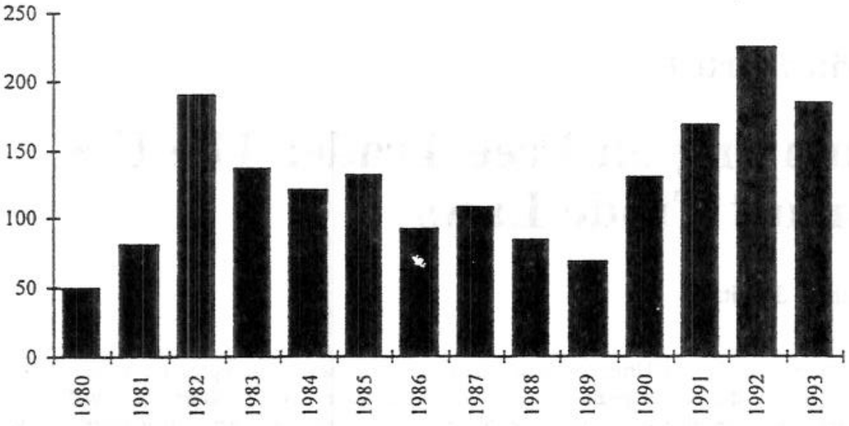


Figure 1. Antidumping Cases Filed (Non-U.S. GATT Signatories)

tion—the culmination of the eighth round of international negotiations under the General Agreement on Tariffs and Trade (GATT). The conclusion of the Uruguay Round seems an appropriate occasion to review the policy goals of our import trade statutes,[3] to assess how well our current laws achieve those objectives, and to explore future courses of action. This paper is therefore organized into four sections: Section 2 discusses the appropriate policy objectives of our import trade laws, section 3 describes the operation of our pre-Uruguay Round statutes, section 4 explores some of the changes effected by the Uruguay Round, and section 5 examines possible directions for future policies.[4]

2. What Are the Proper Policy Objectives of Import Trade Laws?

A well-designed set of import trade laws would address at least three crucial economic policy concerns: dumping, foreign export subsidies, and import surges.[5] This section will discuss such concerns and their impact on national economic welfare.

Dumping

Dumping is conventionally defined as a type of international price discrimination: the sale of goods within the United States at a price lower than in some foreign market.[6] Ironically, viewed from a static perspective, laws proscribing such international price discrimination prevent behavior beneficial to the U.S. and condone behavior detrimental to the U.S. Consider the example of a monopolistic foreign firm. If there are barriers to trade that facilitate price discrimination, our static welfare is maximized if the foreign firm discriminates when our demand is relatively elastic and does not discriminate when our demand is relatively inelastic. (Under conventional assumptions, the price charged to U.S. consumers would be lower than a single pooled price if our demand is more elastic than foreign demand. It would be higher than a pooled price if our demand is less elastic than foreign demand.) But antidumping laws prohibit price discrimination in the first case and not in the second—exactly the opposite of what our optimal policy should be! Even in more complicated static cases (for example, oligopolistic competition) with domestic producers, dumping is still frequently welfare-improving.[7]

---

[3] The unfair import trade laws are a subset of the import trade laws: The import surge law, for example, does not purport to address unfair trade.

[4] In order to focus the discussion, we will concentrate on those laws aimed at restraining unfair import trade; laws that can result in import restraints but that are primarily aimed at removing impediments to exports—such as Section 301—will be discussed only tangentially. For a thorough examination of Section 301, see Bayard and Elliot (1994). We will also exclude laws dealing exclusively with agricultural concerns (e.g., Section 22 of the Agricultural Adjustment Act) or environmental issues (e.g., the Pelly Amendment).

[5] There may be other justifiable policy concerns, like national defense objectives, but they lie outside the purview of this analysis.

[6] For a relatively early analysis of dumping, see Viner's (1923) seminal work. He advocates the definition given in the text.

From a static perspective, then, dumping by foreign firms frequently seems to make us better off. But there are two reasons to be wary of the predictions of the static perspective and thus two legitimate reasons for policymakers to worry about dumping: predatory pricing and new trade theory effects.

Predatory Pricing

Economists have long recognized the possible welfare losses caused by predatory pricing, in which a firm intentionally drives down the market price in order to force other firms out of the market.[8] Compared to the counterfactual of no predatory behavior, consumer surplus is higher during the predation period (since the price is lower) but then lower in the subsequent period (since, if the predation is successful, market power and the price are higher). After discounting and integrating appropriately, it is certainly possible that consumer surplus will be reduced by predation. Does economic theory offer policymakers any guidance in assessing the likelihood of such losses?

The theory of predatory pricing has evolved substantially over the past decade.[9] The older Chicago view is that predatory pricing is unlikely to occur, largely because in most situations it is implausible for firms to expect such a strategy to generate positive (discounted) profits. According to this view, the predation period may have to be relatively prolonged and costly. Entry during the later period, furthermore, is likely to dissipate any incipient rents. The implication is that predatory pricing is rarely a profitable activity and therefore should rarely be observed. New developments in industrial organization theory, however, suggest a somewhat less sanguine perspective: They posit a variety of ways in which predators can prevent entry and thus protect the rents accruing during the second stage. The strategic entry deterrence literature, for example, presents models with irreversible investment, in which firms can deter entry by building excess capacity or investing in advertising. Such activities can be expensive, but they may make predatory pricing a profitable strategy by protecting the second period rents. Alternatively, predatory pricing itself can be a deterrent to entry: Such behavior may create a reputation for craziness and irrational behavior, and this reputation may dissuade potential rivals from entering in the later period. Either way, a predatory pricing strategy may be more attractive to firms than the Chicago view would suggest.

---

[7] See, for example, Brander and Krugman (1983). In general, a domestic welfare gain from foreign dumping may be more likely if we are net importers of the product. In partial equilibrium, the net import assumption ensures that the consumer surplus gain from a lower price dominates the domestic producer surplus loss in the market for the imported good. But matters are more complicated once other markets are included in the analysis (assuming nonseparability of utility across commodities) and in a dynamic context.

[8] Such concerns are only germane in an imperfectly competitive environment: Within a perfectly competitive world, antitrust issues are obviated because free entry obtains by assumption (ensuring zero profits in equilibrium).

[9] For an overview, see, for example, Baker (1989).

If predatory pricing actually occurs, a foreign predator may reduce domestic welfare more than a domestic predator *ceteris paribus* for at least two reasons.[10] First, a predatory strategy will be adopted only if it generates expected discounted profits in excess of the cost of exercising the option to undertake the predation.[11] Thus, monopolization of a domestic market by a foreign producer creates an expected domestic welfare loss both from the deadweight loss (as in purely domestic cases) and from the expected present discounted value of the differential profits accruing from the strategy, which represent an expected transfer of rents from domestic consumers to the foreign producer. Second, labor is usually thought to be relatively mobile within a country—especially in the United States—but relatively immobile between countries. If workers in the firm that is attacked have skills specific to the industry, the transfer of production to the predator firm generates more domestic dislocation (in terms of unemployment mismatch) if the expanded production occurs abroad rather than at home. This seems more likely if the predator is a foreign firm rather than a domestic one.

Predatory dumping may thus be a legitimate policy concern. But it is important to note that dumping, as conventionally defined, is neither necessary nor sufficient for predation in practice. Predation can occur even if the price is higher in the U.S. than abroad (if the U.S. price, for example, would have been even higher, relative to the foreign price, in the absence of the predatory behavior). And price discrimination across national boundaries can be nonpredatory (if, for example, elasticities of demand for a firm's product vary across countries). To foreshadow the discussion in section 3, I should also stress that a U.S. price below the foreign exporter's average cost is neither necessary nor sufficient for predation. Predation is possible even if the price is above the predator's average cost (if, for example, the price is below both the prey's marginal cost and the predator's short-run profit-maximizing price[12]). And prices below average cost—or even below short-run marginal cost—are not *prima facie* evidence of predation (see section 3).

New Trade Theory Effects

The new trade theory suggests that there may be a limited number of cases, involving some sort of nonconvexity and imperfect competition, in which protection of the home market affords a firm first-mover advantages, facilitates dynamic learning effects, or generates other benefits.[13] I should note, however, that it is not obvious that outright protection of the home market, as opposed to subsidization, actually bestows any new trade theory advantages on domestic firms. Under a protectionist policy, the domestic firm obtains a larger share of a smaller market (assuming a downward-sloping domestic demand curve). Whether or not protection produces new trade theory benefits for the domestic firm thus depends on the specific form of the relevant new trade theory effect, the effectiveness of the protection in excluding foreign firms, and product demand parameters (see Brander and Spencer 1984; Krugman 1984).

---

[10] This does not necessarily suggest that policymakers should be disproportionately concerned about international cases, however. The presence of a foreign producer in the market may indicate that the environment is less conducive to a successful predatory strategy, thus invalidating the *ceteris paribus* assumption above, if, for example, it indicates that entry by other foreign producers is relatively easy in the particular market (i.e., because import barriers are low). In other words, international predation may be both more deleterious and less likely.

[11] Predation is an irreversible investment undertaken in an environment of uncertainty and thus has an option value that must be accounted for in the firm's predatory strategy. For an introduction to the options approach to irreversible investment under uncertainty, see Dixit and Pindyck (1994).

[12] In this example, there are efficiency benefits to the predation process (the predator's marginal cost is below the prey's marginal cost, assuming the predator's marginal cost is no higher than her average cost). But these efficiency benefits must be compared to the distortions induced by the predator's increased market power in order to reach a conclusion as to whether the predation increases social welfare.

[13] For basic introductions to the new trade theory, see Krugman (1986) and Helpman and Krugman (1989, Chapter 5).

Protectionism may not produce new trade theory benefits in all industries, but the potential gains in some industries raise complex policy questions. In particular, in a new trade theory world, the strong (conventional) form of predation must be supplemented by a weak form of predation, in which the foreigner's objective is to acquire or appropriate rents from domestic firms rather than to drive them out of business.[14] The weak form of predation suggests that antidumping laws could potentially boost national welfare either by allowing domestic firms to gain new trade theory advantages or by preventing foreign firms from obtaining them.

Foreign Export Subsidies

In the majority of cases, it is in our national interest to welcome subsidized exports from other countries. Such subsidized exports most frequently occur in industries in which we are net importers so, at least from a partial equilibrium analysis of the market for the imported good, the consumer surplus gain from any price decline exceeds the producer surplus loss.[15] Thus, economists often argue that we should be grateful for the gift, perhaps even sending a "thank-you note," as one trade scholar glibly put it (Jackson 1989, p. 251; also see Dixit 1987). But there are two reasons to temper our general enthusiasm for subsidized imports: the new trade theory (again) and global welfare maximization.

New Trade Theory Effects

The new trade theory suggests that there may be some cases in which export subsidization could raise national welfare and reduce welfare in the importing country (Brander and Spencer 1985). In this case, countervailing duty laws (which countervail subsidies that promote exports) can be useful in preventing foreign firms from gaining a first-mover advantage. Similarly, multilateral disciplines and foreign countervailing duty laws can mitigate the incentives for us to try to gain such an advantage.

Global Welfare Effects

Even within the conventional paradigm of perfectly competitive markets, the existence of countervailing duty laws could be beneficial from a global welfare perspective. Such laws attempt to counteract the effects of export subsidies and thus may dissuade misguided policymakers from granting the subsidies in the first place.[16] Put another way, subsidies that distort the trade flows consistent with comparative advantage reduce global welfare (again, assuming a perfectly competitive world). Countervailing duty laws may be useful in discouraging such behavior. The problem is that the U.S. has to threaten to shoot itself in the foot in order to ensure that others do not shoot themselves in their respective feet. We can only hope that the threat is credible enough that we will not have to execute it (and hence ourselves).

---

[14] In a Cournot model, for example, a reduction in one firm's marginal cost can result in an increase in its share of the market, even though the second firm does not vanish. So if the foreign firm's marginal cost is reduced because of new trade theory effects, it can capture a larger share of the rents in the market even if it does not drive the domestic firm out of business.

[15] From a different but equivalent partial equilibrium perspective, our terms of trade improve. In general equilibrium, the impact on social welfare may be more complicated (e.g., if utility functions are not separable across commodities and if cross-elasticities are significant).

[16] The laws may also discourage rent-seeking behavior from firms eager for the subsidies.

A final subtlety concerning countervailing duties is the treatment of externalities. Global welfare may be boosted by government subsidies that encourage activities with positive externalities (e.g., research and development [R&D] and other activities). And to the extent that the externalities cross national boundaries, our national welfare may be raised by foreign government subsidies. So there is some motivation for exempting R&D and other similar subsidies from the countervailing duty laws. At the same time, however, it is often difficult in practice to distinguish a production subsidy from an R&D subsidy. This is the rationale for the treatment of R&D subsidies in the Uruguay Round agreement (see section 4): They are protected from countervailing duties, i.e., are green-lighted, but there are limits on what qualifies as an R&D subsidy.

Import Surges

With large adjustment costs and some form of market failure (for example, capital market imperfections), it may be in the national interest for the government to provide a means of smoothing adjustment to very large but temporary trade shocks, basically as a type of insurance policy, even though such shocks do not represent unfair trade.[17] The problem with this kind of policy is that it is not possible to discern whether a given shock is temporary or permanent until well after the shock has hit. Since policies that impede adjustment to permanent shocks are likely to be welfare-reducing, the optimal market surge policy must compare the (potential) benefits of providing insurance against temporary surges to the (probable) costs of preventing or delaying required adjustments to permanent shocks.

Whether or not the market surge laws provide a direct economic net benefit to the country, they can be supported on political economy grounds. Some level of insurance seems to be necessary to make a free trade regime politically acceptable, and thus a market surge protection law may be required to obtain the overall benefits of a liberal trading regime. The GATT recognizes this reality in Article XIX, which allows safeguards to be imposed if unanticipated increases in imports inflict serious injury on a domestic industry.

3. The Pre-Uruguay Round Trade Laws

The U.S. has an elaborate and comprehensive system of trade laws, many of which are referred to by their statutory numbers. Sections 201 and 406 address market surges,[18] Section 232 regulates imports that threaten national security, Section 301 deals with foreign trade barriers and violations of trade agreements, Section 337 addresses unlawful marketing of imports (such as infringements of patents), Section 701 governs countervailing duties, and Section 731 is the antidumping statute.[19] My focus here will be the antidumping, countervailing duty, and market surge laws. I will also discuss Section 337 briefly.

---

[17] Available derivatives markets are not sufficient to hedge against large import surges.

[18] The import surge laws are often referred to as escape clauses or safeguard statutes. But import surge laws seems to be the most accurate description of their substance.

Antidumping Laws

The current antidumping process has two parts: a dumping component and an injury component. Under the dumping part of the process, the Department of Commerce compares the price in the U.S. to a so-called foreign market value; the dumping margin is then simply the difference between the foreign market value and the U.S. price. Under the injury part, the International Trade Commission (ITC) determines whether the imports are causing or threatening material injury to the U.S. industry. (This injury standard is substantially weaker than in the import surge laws.) Both dumping and injury must be found for final duties to be imposed, and relief must be granted if both findings are positive.

Unfortunately, our antidumping laws no longer have much to do with the prevention of the strong form of predatory pricing.[20] Our first antidumping statute—the Antidumping Act of 1916—was clearly aimed at preventing international predatory pricing.[21] But since the 1921 Antidumping Act, it has not been necessary to demonstrate predatory intent or effect, so any international price discrimination (whether predatory or not) in which the U.S. price is lower than the foreign price has been proscribed.[22]

Since the 1974 trade act, the focus of the antidumping laws has shifted from preventing price discrimination to preventing sales below fully allocated average cost.[23] The mechanism driving this result is buried deep in the trade statutes. The law defines a hierarchy of methods for determining the foreign market value of a good: the home market index (the price at which the good is sold in the exporter's home market), the third country index (the price at which the good is sold in another of the exporter's foreign markets), and, only as a last resort, constructed value (an estimate of the exporter's average cost).[24] This hierarchy seems somewhat reasonable. But the law also stipulates that sales below total cost must be disregarded when calculating the first two indices if such sales have been made "over an extended period of time and in substantial quantities, and are not at prices which permit recovery of all costs within a reasonable period of time in the normal course of trade" (Committee on Ways and Means 1993, p. 388).

---

[19] For some reason, the antidumping and countervailing duty laws are not usually referred to by their section numbers.

[20] In the discussion below, all references to predatory pricing and predation refer to the strong form, that is, to price reductions intended to drive other firms out of the market.

[21] The law included a requirement that the dumping had to "be done with the intent of destroying or injuring an industry in the United States, or of preventing the establishment of an industry in the United States, or of restraining or monopolizing any part of trade and commerce in such articles in the United States" (Committee on Ways and Means 1993, p. 417).

[22] The legislative history of the 1921 Act suggests that, despite the formal language of the law, the prevention of predatory pricing was still its primary objective. As the committee report states, the Act "protects our industries and labor against a now common species of commercial warfare of dumping goods . . . until our industries are destroyed, whereupon the dumping ceases and prices are raised at above former levels to recoup dumping losses" (quoted in Dale 1980, p. 12). See also Nivola (1993, p. 31).

[23] Congressional hearings held in 1968 to consider the Kennedy Round antidumping codes clearly indicate that protagonists then viewed the prevention of price discrimination as the primary objective of the antidumping laws. In his opening statement, for example, Senator Russell Long is quoted as saying, "The Antidumping Act . . . is designed to protect against the rankest kind of commercial injustice—*price discrimination* by foreign competitors . . . " (emphasis added) (Committee on Finance 1968, p. 1). See also Cass and Narkin (1991, pp. 208–212).

[24] The methodology used in cases involving nonmarket economies is different.

In practice, so many observations are excluded because of this provision that foreign market value must often be calculated using the constructed value index.[25]

Prohibiting sales below average total cost precludes many perfectly rational and nonpredatory activities.[26] For example, economists have long recognized that profit-maximizing firms may have sales below average total cost but above average variable cost in the short run. Furthermore, as Spence (1981) and others have shown, learning curves can induce firms to forward price at expected long-run marginal cost rather than short-run marginal cost, so prices below current marginal cost (and current average cost) do not necessarily imply that predatory behavior is occurring.[27] And newer theories suggest other motivations, such as those generated by irreversible investments, for profit-maximizing firms to price below average variable cost over some (perhaps relatively extended) periods.[28]

The pre-Uruguay Round antidumping statutes contained a legion of other serious biases, some of which were partially addressed by the Round (see section 3).[29] For example, individual transactions in the U.S. were compared to averages abroad and negative dumping margins were counted as zero, so any dispersion in the time series of prices produced a positive dumping finding, even if there was no dumping on a transaction-by-transaction basis.[30] In addition, the laws imposed minimum rates for profit (8%) and overhead (10%), regardless of the firm's actual profit or overhead. This tended to produce a higher constructed value, and thus a higher dumping margin, for firms with profit or overhead rates under the minimum. And misuse of the best-information-available process often meant that useful information submitted by foreign exporters was disregarded.[31] These provisions, as well as numerous others, resulted in positive dumping margins when no real dumping had occurred.

---

[25] The Commerce Department's pre-Uruguay Round methodology disregarded all foreign price data and used the constructed value index if more than 90% of foreign sales were deemed to be below cost.

[26] The Areeda–Turner predatory pricing criterion recognizes that sales below average total cost are not necessarily predatory (see Areeda and Turner 1975, pp. 697–733).

[27] See Spence (1981). Spence's model assumes no uncertainty and a zero discount rate. For a discussion of learning curve effects with uncertainty and a positive discount rate, see Dixit and Pindyck (1994, pp. 339–353).

[28] See Dixit and Pindyck (1994, pp. 292–293).

[29] For more detail on many of the biases in the pre-Uruguay Round dumping laws, see the collection of essays in Boltuck and Litan (1991). The collection also includes an essay by an enthusiastic supporter of the statutes (see Stewart 1991, pp. 288–330).

[30] Suppose that sales in both the foreign market and the U.S. market occurred on two dates. On the first date, the good was sold for $10 in the U.S. market and for the equivalent of $10 in the foreign market. On the second date, the good was sold for $5 in the U.S. market and for the equivalent of $5 in the foreign market. Clearly, there was no dumping; the goods were sold for the same price on both occasions. But the transaction-to-average comparison process would nonetheless have produced a positive dumping margin: The average price for foreign sales was $7.50, so the second sale in the United States was dumped, with a margin of $7.50 − $5.00 = $2.50. The first sale in the U.S. had a dumping margin of −$2.50, so that if the two individual margins were averaged, the correct answer (zero) would be obtained. But instead the first dumping margin would have been set to zero, and the final average dumping margin would therefore have been $1.25 (=$2.50/2).

[31] A recent article in *The Washington Post* noted that "the administrative burden simply of furnishing the required information within the required time in the required form to the Department of Commerce has become so overwhelmingly difficult that more and more foreign companies are either unable or unwilling to try . . . [C]ompanies of the stature of Matsushita, SKF, and Toshiba find compliance with the Commerce Department's demands impossible . . ." (Thompson 1994). Not surprisingly, there is some evidence that cases that did not use information from foreign exporters resulted in larger dumping margins than cases that did. Baldwin and Moore (1991, Tables 701, 703) find that the average dumping margin for final determinations using information from foreign firms was 27.9%, whereas the average dumping margin for final determinations that did not use such information was 66.7%. In a pooled regression of all final dumping margins, the dummy variable for the lack of foreign information was positive (0.383), with a t-statistic of 4.19. But it is unclear whether these results are causal or merely an example of selection bias: It could be that the

Table 1. U.S. Antidumping Cases by Industry, 1980–1989

Industry	Number of Cases Filed
Chemicals	58
Food	16
Iron and Steel	201
Machinery	8
Nonferrous Metals	16
Textiles and Apparel	15
Other	137
Total	451

Source: Boltuck and Litan (1991, Table 1-2, p. 3).

Given the focus on costs and the biases in the process, a dumping finding was almost a foregone conclusion, especially in the capital-intensive, highly cyclical industries that account for the majority of dumping cases (see Table 1). As Lester Thurow has noted, if "the law were applied to domestic firms, eighteen out of the top twenty firms in the Fortune 500 would have been found guilty of dumping in 1982" (Thurow 1985, p. 359). From this perspective, it is perhaps not surprising that a positive dumping margin was found in over 80% of all cases that reached a final determination in the early and mid-1980s.[32] So the dumping test often did not function as a binding constraint.[33]

Recent research has substantiated other distortions caused by the antidumping process, many of which are likely to continue in the post-Uruguay Round period. As one perhaps predictable example, empirical research has found that firms may not have to wait for a positive final determination in order to benefit from a dumping suit. Staiger and Wolak (1994b) report that "petitioning firms may enjoy import relief during the investigation period that amounts to about half of what they might expect from a positive final determination and duty imposition" (Staiger and Wolak 1994b, p. 101). There are several possible explanations for this result, including the retroactive calculation and imposition of duties at the end of the investigation (so that exporters and importers have an incentive to raise prices immediately once an investigation has begun) and the adverse effect of an investigation on competitive pressures within an industry. Whatever the explanation, the result has been that the antidumping laws are becoming increasingly attractive to firms seeking protection against imports—and not necessarily unfair imports.[34]

Because positive dumping margins are almost always found and because any significant fall in sales is sufficient for a positive injury finding (even if most of the fall in sales is caused by other factors), the antidumping laws are now effectively acting like market surge statutes,

---

foreign firms most unfamiliar with the antidumping process or with the least incentive to comply with that process were also the ones with the largest dumping margins.

[32] Murray (1991, Table 2-3, p. 46). The figure applies to cases initiated after 1979 and concluded before 1986. It excludes cases that were terminated or suspended. According to a press report, positive determinations have been even more likely recently: An article in *Forbes* magazine alleges that dumping was found in 97% of the cases decided between 1988 and 1992 (see Frum 1992).

[33] We should note, however, a possible selection bias: The cases in which dumping would not be found may not be observed as frequently because petitioners can predict the outcome and therefore a disproportionate number of such cases are not filed. This criticism does not seem relevant to Lester Thurow's observation.

[34] Between 1963 and 1979, U.S. firms filed 532 antidumping and countervailing duty cases and 75 import surge cases. Between 1980 and 1988, they filed over 700 antidumping and countervailing duty cases and just 19 import surge cases (Hansen and Prusa 1993, p. 35).

especially for moribund industries. Indeed, the laws are now sometimes referred to as the baby 201 laws.[35] This baby 201 tendency is the direct result of provisions within the import surge laws (like Presidential discretion, a higher injury standard, and the limited duration of remedies) that are intended to avert their misuse. In effect, then, the antidumping laws are functioning as an ersatz market surge system without the restrictions intended to prevent abuse.

According to many analysts, the antidumping laws are indeed being abused. This is perhaps most apparent in the filing of harassment cases intended to impose purely temporary trade restraints and legal costs on foreign exporters (Staiger and Wolak 1994a). Such cases are particularly effective because of the asymmetries in legal costs borne by domestic plaintiffs and foreign defendants. The current chairman of the ITC has been quoted as saying that he "can identify areas where there are abuses . . . petitioners which use the trade laws purely as aggressive tools, rather than as a legitimate response to unfair trade practices."[36] Staiger and Wolak (1994b, p. 100) estimate that 10 of the 338 manufactured goods cases adjudicated between 1980 and 1985 were harassment or process cases—a small number, to be sure, but nonetheless worrisome.

In summary, the antidumping laws no longer have very much to do with the prevention of predatory pricing. Indeed, a recent OECD study examining U.S. dumping cases between 1979 and 1989 concluded that conditions were not even conducive to predatory pricing in over 85% of the cases that found dumping or that were terminated or suspended before a final determination.[37] It is perhaps worth stressing that the prevention of predatory pricing is a more restrictive standard than the prevention of price discrimination, which is itself a more narrow concept than preventing below-cost sales. And we do not even conduct our price-to-cost comparisons in an entirely disinterested manner and thus are at least three steps away from a purely predatory pricing rule.

As for the possible benefits of an antidumping law within a new trade theory environment, Boltuck and Litan (1991, p. 11) note that "there is no evidence that the antidumping law in the United States or in any other country is either designed or administered in a manner to ensure that its application is limited to such instances."

Countervailing Duty Laws

The first U.S. statute to address unfair international trade was an 1897 countervailing duty law addressing export subsidies (Committee on Ways and Means 1993, p. 53). This statute remained largely unchanged between 1922 (when foreign domestic subsidies were included) and 1979 (when the U.S. implemented the GATT Subsidies Code, which had been negotiated as part of the Tokyo Round).[38] The 1984 and 1988 trade acts made some other changes to the countervailing duty laws.

Under the countervailing duty law, the Department of Commerce determines whether imports into the U.S. are being subsidized, and the ITC determines whether material injury results from such subsidized imports. (The injury test is applied only in cases against GATT signatories.) As in antidumping, the procedure tends to produce biased margins, and the injury standard is low. Since it is relatively easy under these laws to show that a foreign firm has been subsidized in some way, the countervailing duty laws have become a popular protectionist sibling to the antidumping laws.

---

[35] The sobriquet applies to the countervailing duty laws also (Boltuck and Litan 1991, p. 11).

[36] See *Financial Times* (10 August 1994).

[37] See OECD Working Party No. 1 (1994).

[38] The most important reform in 1979 was the introduction of a mandatory injury test for all signatories to the Code.

One contentious issue in countervailing duty law is the treatment of subsidies to firms that are subsequently privatized through auctions. It would seem to most economists that any subsidy granted prior to privatization would be capitalized into the privatization price. So the government would have been paid back, and the subsidy should not be countervailable, after privatization. Indeed, the Court of International Trade recently advanced such an argument.[39] This issue is of particular importance for the economies in transition and will be discussed further below.

Import Surge Laws

Section 201 is the primary U.S. market surge law, and it is generally consistent with the stringent requirements of the GATT safeguards clause (Article XIX). In particular, Section 201 requires that the ITC determine whether "an article is being imported into the United States in such increased quantities as to be a substantial cause of serious injury . . . to the domestic industry . . . " (Committee on Ways and Means 1993, p. 452). If the ITC concludes that these requirements are satisfied, it then submits a suggested remedy to the President. The President has the discretion to approve the ITC's suggestion, to adopt a different remedy, or to do nothing at all. In making this decision, the President is instructed to evaluate whether a proposed action provides "greater economic and social benefits than costs." If a remedy is applied, it must be applied against imports from all trading partners (i.e., on a most-favored nation basis) and must be temporary (which is defined as anything lasting no more than eight years) (Committee on Ways and Means 1993, p. 464; Hansen and Prusa 1993, p. 38). Since it is much easier to obtain relief under the antidumping and countervailing duty statutes, Section 201 is now rarely used. No petitions, for example, were filed during the whole of 1993 (Office of the United States Trade Representative 1994, p. 105).

Section 337

Section 337 primarily addresses the infringement of patents, copyrights, trademarks, or mask works in the importation of goods into the United States (Committee on Ways and Means 1993, p. 444). Under this statute, the ITC investigates whether a violation has occurred and, if it reaches a positive determination, issues a remedial order to stop the imports, an order that the President can overrule. In 1989, a GATT panel concluded that parts of the Section 337 process violated the principle of national treatment because imported goods were subjected to a different legal system (distinct from the corresponding intellectual property rights statutes applying to domestically produced goods) solely on the grounds of their country of origin.[40]

Nonmarket Economies

Finally, the import trade statutes have a distinct set of procedures for so-called nonmarket economies (NMEs) like the countries of the former Soviet Union. The antidumping laws for

---

[39] The case involved the U.K.'s United Engineering Steels, a joint venture between British Steel and Guest, Keen, and Nettlefolds. The Court overturned Commerce's decision to countervail subsidies granted prior to privatization (see *Financial Times* 15 June 1994). This issue is of crucial importance in assessing the benefits of the transfer of non-market economies to market economy status. See discussion in the text that follows.

[40] See Office of the United States Trade Representative (1994, p. 104) and Jackson (1989, p. 192).

NMEs are even more biased than those for market economies, and the NME market surge statute (Section 406) has a lower injury standard than Section 201.

NME dumping cases are determined using the surrogate country constructed cost measurement, in which factor costs from one or more surrogate market economies are combined with factor proportions from the NME to construct an average cost measure. The law stipulates that in choosing the surrogate country, Commerce should, "to the extent possible," rely on "market economy countries that are at a level of economic development comparable to that of the nonmarket economy country, and significant producers of comparable merchandise" (Committee on Ways and Means 1993, p. 389). Even a disinterested application of this methodology would be likely to produce dubious results. But the application of this statute has hardly been disinterested, and the results have often been farcical: Countries like the U.K. and Canada have been used as surrogates for China and Russia (Cass and Narkin 1991, p. 216). In a famous case involving Polish golf carts, Canada was chosen as a surrogate for Poland. But Canada had stopped producing golf carts before the investigation was completed. So after selecting Canada, the administering authority had to determine how much Polish golf carts would have cost in Canada if Canada had produced golf carts.[41]

Nonmarket economies also have unique status under the countervailing duty and market surge laws. They are currently exempt from countervailing duty cases because the courts ruled—in the Georgetown Steel case—that it was impossible to determine the level of subsidies in the distorted price environment that obtains in NMEs. But it is unclear whether subsidies granted while a country holds NME status are countervailable once the country graduates to market economy status. If they are, this would represent a substantial disincentive to moving to market economy status.

The market surge statute that applies exclusively to the NMEs, Section 406, is broadly similar to Section 201. But Section 406 has a lower injury standard (the surge in imports must be a "significant cause of material injury" rather than a "substantial cause of serious injury"), and remedies are applied only against imports from the NMEs, not against all countries (as under Section 201).

4. The Uruguay Round

Some of the most contentious battles of the Uruguay Round negotiations were waged over the unfair trade provisions. This section will briefly describe the relevant features of the Uruguay Round agreement.[42]

In antidumping, the Uruguay Round agreement generally moves world antidumping systems closer to U.S. procedures, though there are also some provisions that require changes in U.S. practice. Table 2 describes the more important changes in the U.S. laws. Examples of marginal improvements in the laws include the stipulations that profit and overhead calculations

---

[41] Technically, this case was adjudicated using a surrogate country price measure: Canada's golf cart price, rather than factor costs, was used to construct the foreign market value. But since Canada did not produce golf carts, the price had to be constructed. In effect, then, the only difference between the surrogate country price measure and the surrogate country constructed cost measure in this case was that Canada's factor proportions were used instead of Poland's. In an administrative review, the surrogate country was switched to Spain.

[42] For an overview of the results of the Uruguay Round and an introduction to the operations of the World Trade Organization, see Jackson (1994).

Table 2. The Uruguay Round and the U.S. Dumping Laws

Issue	Description of Issue	Uruguay Round Changes	Other Comments
Sunset Reviews	Dumping duties are intermittently reviewed. Nevertheless, many are perpetuated for extremely long periods of time. A sunset review establishes the presumption of a limited duration for duties.	Under previous statutes, duties had no definite duration. Under the new legislation, duties are terminated five years after imposition unless termination would be "likely to lead to continuation or recurrence of material injury within a reasonably foreseeable time."	There are some lingering concerns over the standards to be applied during the sunset review (e.g., the definitions of likely and reasonably foreseeable). The procedure applied to existing duties (i.e., the timing and ordering of their reviews) is also the subject of much interest to affected parties.
Averaging	Dumping margins could be calculated by matching individual transactions in each market (a transaction-to-transaction basis), by comparing averages in each market (an average-to-average basis), or by comparing individual transactions in the home market to an average for the foreign market (a transaction-to-average basis). This is the averaging issue.	Under previous statutes, comparisons of individual prices in the U.S. to average prices abroad biased dumping margins upwards. Under the new legislation, price comparisons of home market value and foreign market value will be conducted on an average-to-average or transaction-to-transaction basis in investigations.	The change applies only to investigations. The old (biased) transaction-to-average methodology will continue to be used in reviews of previously imposed margins. There is also an exemption in the statutes permitting the biased methodology to be employed in investigations under exceptional conditions.

must use actual data whenever possible (although some mischief is still possible when transactions are judged to be outside the ordinary course of trade, which is potentially a major loophole); that price comparisons must be on a transaction-to-transaction or average-to-average basis (although only in investigations, not reviews, and there are some exceptions even for investigations); that orders must be terminated after five years (unless such a revocation is

Table 2. Continued

Issue	Description of Issue	Uruguay Round Changes	Other Comments
Level of Price Comparison	Comparisons between prices of domestic sales and prices of foreign sales are skewed if the sales occur at different levels of trade (e.g., wholesale vs. retail). Therefore, Commerce adjusts the prices to a common level of trade.	The implementing legislation may have introduced some new biases into this process. For example, under the new methodology, indirect selling expenses and profits will often be subtracted from the U.S. price but not from the equivalent foreign price, thereby boosting the dumping margin.	
Calculation of Profits and Overhead in Constructed Value	When Commerce uses the constructed cost index as its measure of foreign market value, it adds overhead and a profit margin to its estimate of operating costs.	Under the previous statutes, Commerce imposed a minimum profit margin of 8% (of total costs) and a minimum overhead rate of 10% (of operating costs). The new legislation directs Commerce to use actual data (from the firm or industry) to calculate these margins, regardless of whether or not they are below the previous artificial minimums.	The profit margin will be calculated only on sales "in the ordinary course of trade." This caveat could permit Commerce to disregard many sales below average total cost, thereby inflating the profit margin and the dumping margin.

"likely to lead to continuation or recurrence" of injury); and that evidence submitted by foreign exporters must be accepted and used (subject to some relatively weak requirements). While such changes represent a significant improvement in the law and do address some of the more egregious biases in the previous system, they nonetheless perpetuate the fundamental structure of the antidumping process, with its emphasis on preventing sales below average total cost.

The Uruguay Round agreement also addresses subsidies, which it divides into three categories: those that are prohibited (or so-called red light subsidies), those that are permitted unless

Table 2. Continued

Issue	Description of Issue	Uruguay Round Changes	Other Comments
Startup Costs in Constructed Value	Firms often incur large startup costs in a new operation. Such costs should be amortized over the life of the project. The issue is therefore how these costs should be treated in calculating the constructed cost index.	Under the new legislation, startup operations qualify for an adjustment to their production costs. Commerce is directed to substitute "the unit production costs incurred with respect to the merchandise at the end of the startup period for the unit production costs incurred during the startup period."	Startup costs are crucial in many industries, especially capital-intensive ones.
Best Information Available	If exporting firms do not submit adequate information, Commerce bases its conclusions on the "best information available." Often this "best information available" is that submitted by domestic complainants.	Under previous practice, information requirements were difficult to fulfill and information from partially incomplete submissions was often completely disregarded. The new process addresses these shortcomings by easing submission requirements (especially for small firms) and ensuring that any relevant information, even from partially incomplete submissions, is used.	There are lingering questions about the process for verifying and corroborating the information submitted by parties to the case.

they cause adverse trade effects (yellow light), and those that are permitted regardless of trade effects (green light). Red-light subsidies include export subsidies (both *de jure* and *de facto*) and subsidies contingent on local content requirements. Yellow-light subsidies are those not

Table 2. Continued

Issue	Description of Issue	Uruguay Round Changes	Other Comments
De minimis Margins	If the calculated dumping margin is below some *de minimis* threshold, Commerce concludes that no dumping has occurred.	The new statutes impose a *de minimis* margin of 2% in investigations. The previous 0.5% margin is maintained for reviews.	
Negligible Imports	If the dumped imports are insignificant, ITC concludes that no injury has occurred.	Under previous statutes, ITC decided whether the imports were insignificant based on their volume and market share. The new statutes introduce an import share threshold: If imports from any particular country constitute less than 3% of total U.S. imports, they are deemed negligible and the dumping case is dismissed.	There is an exception to the 3% rule if dumping cases are filed against exporters from many countries that collectively account for more than 7% of U.S. imports (even if each one individually accounts for under 3% of U.S. imports). In any case, import shares above 3% are not necessarily indicative of a substantial U.S. presence if total imports are an insignificant share of the U.S. market.

[a] For completeness, an average-to-transaction basis is also possible (in which the average at home is compared to individual transactions abroad). But this is problematic conceptually: For example, if a single sale in the foreign market occurs at a price higher than the U.S. average, are all the U.S. sales dumped?? In any case, such a procedure is not used in practice.

[b] The Uruguay Round implementing legislation has changed several parts of the terminology used in the U.S. statutes in order to bring them into conformance with the GATT 1994. In particular, "foreign market value" is now called "normal value."

otherwise prohibited (under the red-light test) nor protected (under the green-light test) and are actionable only if they cause injury or serious prejudice to another country or somehow impair other trade provisions. Green-light subsidies include those not specific to particular enterprises or industries (that is, those that do not disproportionately benefit particular sectors), R&D subsidies (as long as the subsidies cover no more than 75% of the costs of industrial research or 50% of the costs of precompetitive development activity), regional development subsidies (as long as they are granted to a region considered to be disadvantaged on the basis of certain given objective criteria), and environmental subsidies (as long as they are nonrecurring, limited to

20% of environmental adaptation costs, and available to all relevant firms).[43] Subsidies found to contravene these rules can be disciplined by the new World Trade Organization (WTO) or countervailed by member countries.

Perhaps most importantly, the Uruguay Round agreement institutes a much stronger dispute settlement procedure under the auspices of the WTO. Under the new process, a panel decision will be adopted unless there is a unanimous agreement not to adopt it (under the previous GATT dispute settlement process, a panel decision was only adopted if there was a unanimous agreement to adopt it). An approved WTO panel decision against a particular country will require either a change in the relevant practice or the granting of compensation to other WTO members. (Compensation usually means that the other WTO nations are allowed to withdraw previous tariff concessions, providing one of the few types of compensation that can harm both the donor and the recipient.)

5. The Future: How Should We Improve Our Import Trade Laws?

The Uruguay Round brings our trade laws slightly closer to the principles adduced in section 2. But there is still an obvious gap. In policy interventions, as in statistical testing, there are both Type I errors (intervening when we shouldn't) and Type II errors (not intervening when we should). Our import trade laws seem to have an unnecessarily strong bias toward Type I errors and thus are inconsistent with President Clinton's often-quoted exhortation that "we must compete, not retreat." It is generally acknowledged that the laws need to be reformed. Even the chairman of the ITC admits that "we all know these laws can be improved."[44]

Because other countries are learning to mimic U.S. behavior, a more liberal set of import trade laws would bring indirect efficiency benefits—through less restrictive laws abroad—in addition to the direct effects (Congressional Budget Office 1994, p. 7). Indeed, U.S. exporters already face more antidumping cases than exporters from any other country (see Figure 2),[45] and any biases in our laws that protect our import-competing firms are likely to redound to the detriment of our exporters.[46] And our most competitive exporters, the ones most likely to price below some given norm, are the ones most likely to be subjected to antidumping duties. Perpetuating unfair trade laws that are themselves unfair thus imposes substantial burdens on our consumers and on our most efficient exporters while protecting our least efficient import-competing firms.

Even if the import trade laws are not liberalized, there is still a variety of ways to improve them. Our antidumping and countervailing duty laws are now used for a variety of purposes

---

[43] The traffic light rules are somewhat different for less developed countries.

[44] See *Financial Times* (10 August 1994).

[45] Figure 2 depicts the number of antidumping cases filed against exporters from the listed countries and reported to the GATT Committee on Antidumping Practices between January 1989 and June 1993. Only signatories to the antidumping code within the GATT are required to file such reports, and many countries (including many GATT Contracting Parties) are not signatories. In addition, some signatories (e.g., the European Union) do not report some cases filed against nonsignatories. The figures presented thus represent a lower bound on the total number of antidumping cases against the given countries. It is possible that the rankings for the total number of antidumping cases, filed by GATT and non-GATT signatories, would differ from the rankings given here. Source: Compilations based on GATT reports.

[46] The biases in our own laws vitiate our position in criticizing similar biases in other nations' laws. It seems highly unlikely that the U.S. government would file complaints with the WTO against other countries for effectively mimicking our laws.

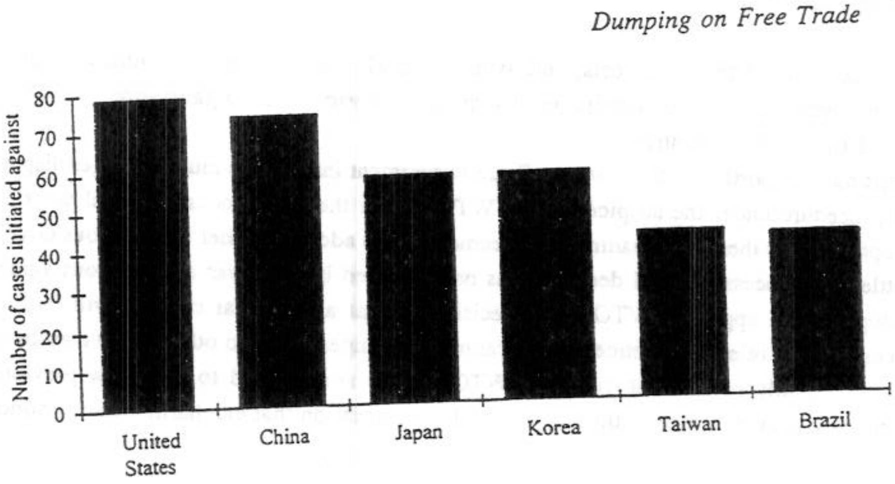


Figure 2. Targets of Antidumping Cases, January 1989–June 1993

for which they were not intended, while spurious accusations of unfairness obfuscate the real policy issues involved. Meanwhile, the domain of the laws is becoming ever more expansive. Trade attorneys, for example, sometimes argue that the antidumping laws must be used to combat foreign monopolies that restrict market access to U.S. firms, even though Section 301 includes an explicit provision to address such problems.[47] Basically, dumping has come to be defined as anything that can produce a positive dumping margin under the extant antidumping statutes. So even if changes are made only on a restrictiveness neutral basis, our system of trade laws could be more transparent and efficient if it functioned as it ostensibly should: with import surges handled under the import surge law, foreign anticompetitive practices handled under Section 301 (or in multilateral form), etc.

What specific reforms would improve the system? Numerous scholars of the laws have put forward suggestions that range from the incremental to the momentous.[48] Below I delineate one possible combination of these suggestions that would simplify the structure of the laws while moving us closer to the policy objectives advanced in section 2. As will become apparent, we do not necessarily abide by the restrictiveness neutral constraint.

Before turning to the specific proposals, I note three general points. First, a crucial issue in considering reform of the import trade laws is the relevance of the new trade theory effects mentioned in section 2. Within the economics community, there continue to be reservations, mainly based on political economy arguments, about the practical benefits of policies predicated on the new trade theory.[49] And any import trade law designed to capture the benefits suggested by the new trade theory seems prone to serious Type I errors: Numerous trade restrictions would undoubtedly be imposed when there are no rents to be captured or protected by the nation as a whole. My conclusion is that the precepts of the new trade theory are an important caveat and may be crucial in a limited number of specific cases, but attempts to exploit such effects within the import trade laws are likely to be both ineffective and costly. Second, any set of trade laws, regardless of how well the laws are formulated, is likely to induce wasteful rent-seeking behavior. Statutes offering even the possibility of protection inevitably engender rent-seeking activities that are both direct (e.g., lobbying) and indirect (e.g., manipulating output in order to make a positive injury finding more likely). Leidy (1994) describes the various ways in which such behavior manifests itself. We can only hope to minimize, not eliminate, such effects. Finally, although I do not explore the issue here, the inclusion of further improvements to our trade adjustment assistance programs could increase the political viability of any reform package while possibly yielding additional economic benefits.

---

[47] See Jackson (1989, p. 212). Section 301 includes in its definition of "unreasonable acts" the "toleration by a foreign government of systematic anticompetitive activities by private firms or among private firms in the foreign country that have the effect of restricting . . . access of United States goods . . . ."

[48] See, for example, Jackson (1989, pp. 242–244), Boltuck and Litan (1991), Trebilcock (1990), Hart (1990), and especially Finger (1993, Chapter 4).

[49] See, for example, Krugman (1987, 1994). Francois (1992, p. 185) concludes that, even in the presence of increasing returns, "on net optimal commercial policy may still best be approximated by free trade."

Antidumping

The most extreme reform to the antidumping laws would be to merge them with our domestic predatory pricing laws. If predatory pricing is the primary policy concern raised by dumping (that is, if the weak form of predation is, in practice, unimportant), this suggestion seems to have substantial merit.

Many of the problems involved in designing a good antidumping law are also faced in designing a good domestic competition law, including the difficulties of measuring marginal cost (as opposed to average cost), the treatment of costs for new operations (startup costs), the nexus between market barriers and the probability of successful predation, and the definition of the relevant market. It seems foolish and inefficient to struggle with these questions twice, once in our domestic antitrust laws and then again in the international dumping laws. Furthermore, for reasons of fairness and simplicity, a single set of predatory pricing laws (for both domestic and international cases) seems preferable. Such a change in our laws would mirror the increasing harmonization and unification of competition policies internationally. In customs unions and free trade areas like the European Union and the Australia–New Zealand Free Trade Area, for example, antidumping laws no longer apply to trade between member countries. Instead, competition law standards are used (Council of Economic Advisers 1994, p. 240).

To be fair, there are some considerations that are specific to the international sphere. Additional complications in international cases include the problems of converting different currencies into a common one (i.e., choosing which exchange rates to use),[50] the rents accruing to foreign producers, the higher adjustment costs caused by the relative immobility of labor across national boundaries, and the possible involvement of sovereign foreign governments. But these difficulties are not intractable.[51]

Other, less severe, reforms to the antidumping laws are also possible. For example, it would not be difficult to remove or attenuate some of the more egregious procedural problems (for example, the focus on price-to-cost comparisons) simply by passing new laws or reinterpreting extant laws. Discretion could be incorporated into the process by allowing the President to veto dumping duties (although I note that the current quasi-judicial system was developed in part to protect the President from the political pressures inevitably associated with any decision on dumping duties). A more comprehensive definition of injury is an especially promising possibility. Finger (1993), for example, has suggested that the injury test should scrutinize the effect on the national economy rather than the petitioning industry. "Antidumping would thus become public policy rather than private policy" (Finger 1993, p. 70).

---

[50] This problem can actually arise in purely domestic antitrust cases, albeit in less severe form, if inputs are imported.

[51] We should note that merging the antidumping laws with the domestic competition laws would probably reduce rent-seeking activities but would not eliminate them: The antitrust laws have also induced much rent-seeking behavior. [illegible] provides an interesting set of examples, including the London Fuller's case of 1298.

Countervailing Duty Laws

In principle, such laws can be beneficial if they are carefully crafted and implemented. But they must no longer be used as a form of protection against market surges. A stricter injury test and a higher standard of proof for establishing the existence of subsidies would be helpful in this regard. Francois, Palmeter, and Anspacher (1991) present a detailed set of the "more obvious reforms that are needed" in the countervailing duty procedures, including the elimination of hypocritical cases (that is, cases against practices also undertaken by the U.S. government), changes in the methodology used to calculate subsidies, and a variety of other specific recommendations.[52] The new World Trade Organization can play a useful role in promoting such reforms.

Import Surge Laws

Protection against market surges may be a legitimate feature of our trade law system, but such protection should probably not be conducted furtively through laws intended for other purposes. Most analysts agree that any surge protection should be transparent, undertaken only in the national economic interest, and betray no connotations of unfair trade. Section 201 seems largely to fulfill these requirements, although one critic has suggested some improvements (the title of his essay, "Throwing Deep: Trade Remedy Laws in a First-Best World," may betray the standard to which he is holding the surge law) (Trebilcock 1990, p. 249–250). Given the existing distortions in the import trade law system, the more urgent reforms surely involve the antidumping and countervailing duty laws.

Section 337

A separate set of laws to protect intellectual property rights violations during the importing process seems superfluous, at least to naive economists like ourselves. Can't all complaints about intellectual property rights be channeled through the corresponding domestic laws? In addition to parsimony and clarity, this would ensure that we are in full compliance with the 1989 GATT ruling.[53]

Nonmarket Economies

The nonmarket economy procedures are widely recognized as severely biased and difficult to defend, especially since the *sui generis* nature of the economies in transition makes comparisons with comparable market economies extraordinarily problematic. Both benevolence and hard-headed national interest, moreover, suggest that the reform process in the former Soviet Union and Eastern Europe should be supported. Open markets—"trade not aid"—seem the most auspicious and mutually beneficial means of providing such support. As Jeffrey Garten, the Undersecretary of Commerce for International Trade, has noted, "At present, most of the economies in transition are still considered non-market economies under the U.S. antidumping law. The United States believes that increased international trade is their best opportunity for development" (Garten 1994, p. 27–28).

---

[52] See Francois, Palmeter, and Anspacher (1991, pp. 130–133).

[53] According to officials in the Office of the U.S. Trade Representative, the Uruguay Round implementing legislation brings Section 337 into compliance with the GATT ruling. But it is not clear that a WTO panel would concur.

One comprehensive reform to the nonmarket economy import trade laws was proposed by the Clinton administration in the spring of 1994. Under the proposal, which is still under consideration, trade complaints against the economies in transition, primarily the states of the former Soviet Union and Eastern Europe, would be channeled through a new import surge law. The antidumping, countervailing duty, and Section 406 processes would be temporarily suspended for these countries. The injury standard under this new law would be higher than in the antidumping code but lower than in Section 201. The ITC would evaluate injury claims and submit recommended remedies to the President. A positive injury finding would require some corrective action by the President (albeit not necessarily the one suggested by the ITC). This proposal seems to represent a substantial improvement over current practice.

6. Conclusion

Abuse of import trade laws represents one of the most ominous threats to a liberal international trading regime. There is a coherent argument for the existence of an extremely circumscribed set of such laws, but our statutes seem to have become divorced from such principles. The Uruguay Round represents a step in the right direction, but only a small step. Economists can play a useful role by continuing to point out the weaknesses in our import trade laws to policymakers and the public, while also proposing and debating possible reforms to improve the functioning of the system.

References

Areeda, P., and D. Turner. 1975. Predatory pricing and related practices under Section 2 of the Sherman Act. *Harvard Law Review* 88:697–733.

Baker, J. 1989. Recent developments in economics that challenge Chicago school views. *Antitrust Law Journal* 58:645–55.

Baldwin, R., and M. Moore. 1991. Political aspects of the administration of the trade remedy laws. In *Down in the dumps: Administration of the U.S. unfair trade laws,* edited by R. Boltuck and R. Litan. Washington, DC: Brookings Institution Press, pp. 253–87.

Baumol, W. 1993. *Entrepreneurship, management, and the structure of payoffs.* Cambridge, MA: MIT Press.

Bayard, T., and K. Elliot. 1994. *Reciprocity and retaliation in U.S. trade policy.* Washington, DC: Institute for International Economics.

Boltuck, R., and Litan, R. 1991. *Down in the dumps: Administration of the U.S. unfair trade laws.* Washington: Brookings Institution Press.

Brander, J., and P. Krugman. 1983. A 'reciprocal dumping' model of international trade. *Journal of International Economics* 15:313–21.

Brander, J., and B. Spencer. 1984. Tariff protection and imperfect competition. In *Monopolistic competition and international trade,* edited by H. Kierzkowski. Oxford: Clarendon Press, pp. 194–206.

Brander, J., and B. Spencer. 1985. Export subsidies and market share rivalry. *Journal of International Economics* 18: 83–100.

Cass, R., and S. Narkin. 1991. Antidumping and countervailing duty law: The United States and the GATT. In *Down in the dumps: Administration of the U.S. unfair trade laws,* edited by R. Boltuck and R. Litan. Washington, DC: Brookings Institution Press, pp. 208–12.

Committee on Finance. 1968. *U.S. Senate international antidumping code.* Ninetieth Congress, Second Session, June 27.

Committee on Ways and Means. 1993. *Overview and compilation of U.S. trade statutes.* Washington, DC: Government Printing Office.

Congressional Budget Office. 1994. *A review of U.S. antidumping and countervailing duty law and policy.* Washington, DC: Congressional Budget Office.

Council of Economic Advisers. 1994. *Economic report of the President 1994.* Washington, DC: Government Printing Office.

Dale, R. 1980. *Anti-dumping law in a liberal trade order.* New York: St. Martin's Press.

Dixit, A. 1987. How should the United States respond to other countries' trade policies? In *U.S. trade policies in a changing world economy,* edited by R. Stern. Cambridge, MA: MIT Press, pp. 245–82.

Dixit, A., and R. Pindyck. 1994. *Investment under uncertainty.* Princeton, NJ: Princeton University Press.

Financial Times, 14 June 1994. Steel duties rejected. p. 4.

Financial Times, 10 August 1994. US trade sleuth in search of fairness. p. 4.

Finger, J. M. 1993. *Antidumping: How it works and who gets hurt.* Ann Arbor: University of Michigan Press.

Francois, J. 1992. Optimal commercial policy with international returns to scale. *Canadian Journal of Economics* 25: 184–95.

Francois, J., N. D. Palmeter, and J. Anspacher. 1991. Conceptual and procedural biases in the administration of the countervailing duty law. In *Down in the dumps: Administration of the U.S. unfair trade laws,* edited by R. Boltuck and R. Litan. Washington, DC: Brookings Institution Press.

Frum, David. 1992. Dump it. *Forbes,* 28 September, p. 64.

Garten, J. 1994. *The United States in a changing world economy: Recent speeches.* Washington, DC: U.S. Department of Commerce.

Hansen, W., and T. Prusa. 1993. Does administrative protection protect? A reexamination of the U.S. Title VII and escape clause statutes. *Cato Review of Business and Government Regulation* Winter:35–6.

Hart, M. 1990. Idealism versus pragmatism: Policy and the academic analyst. In *Fair exchange: Reforming the trade remedy laws,* edited by M. Trebilcock and R. York. Toronto: CD Howe Institute.

Helpman, E., and P. Krugman. 1989. *Trade policy and market structure.* Cambridge, MA: MIT Press.

Jackson, J. 1989. *The world trading system: Law and policy of international economic relations.* Cambridge, MA: MIT Press.

Jackson, J. 1994. Managing the trading system: The World Trade Organization and the post-Uruguay Round GATT agenda. In *Managing the world economy: Fifty years after Bretton Woods,* edited by P. Kenen. Washington, DC: Institute for International Economics.

Krugman, P. 1984. Import protection as export promotion: International competition in the presence of oligopoly and economies of scale. In *Monompolistic competition and international trade,* edited by H. Kierzkowski. Oxford: Clarendon Press.

Krugman, P. 1986. *Strategic trade policy and the new international economics.* Cambridge, MA: MIT Press.

Krugman, P. 1987. Is free trade passe? *Journal of Economic Perspectives* 1:131–44.

Krugman, P. 1994. Empirical evidence of the new trade theories: The current state of play. In *New trade theories: A look at the empirical evidence.* Centre for Economic Policy Research.

Leidy, M. 1994. Trade policy and indirect rent seeking: A synthesis of recent work. *Economics and Politics* 6:97–115.

Murray, T. 1991. The administration of the antidumping duty law by the Department of Commerce. In *Down in the dumps: Administration of the U.S. unfair trade laws,* edited by R. Boltuck and R. Litan. Washington, DC: Brookings Institution Press.

Nivola, P. 1993. *Regulating unfair trade.* Washington, DC: Brookings Institution Press.

OECD Working Party No. 1. 1994. Chapter 2: Census and analysis of antidumping actions in the United States. In *Working Party No. 1 on competition and international trade: Antidumping and competition policy,* draft (6/4/94). Washington, DC: OECD.

Office of the United States Trade Representative. 1994. *Trade policy agenda and 1993 annual report of the President of the United States on the Trade Agreements Program.* Washington, DC: U.S. Trade Representative's Office.

Spence, M. 1981. The learning curve and competition. *Bell Journal of Economics* 12:49–70.

Staiger, R., and F. Wolak. 1994a. The determinants and impacts of antidumping suit petitions in the United States: An industry level analysis. In *New trade theories: A look at the empirical evidence.* Centre for Economic Policy Research.

Staiger, R., and F. Wolak. 1994b Measuring industry-specific protection: Antidumping in the United States. In *Brookings papers: Microeconomics 1994.* Washington, DC: Brookings Institution Press, pp. 51–118.

Stewart, T. 1991. Administration of the antidumping laws: A different perspective. In *Down in the dumps: Administration of the U.S. unfair trade laws,* edited by R. Boltuck and R. Litan. Washington, DC: Brookings Institution Press, pp. 288–330.

Thompson, T. 1994. Dumping on trade. *The Washington Post,* 18 July, p. A19.

Thurow, L. 1985. *The zero sum solution: Building a world-class American economy.* New York: Simon and Schuster.

Trebilcock, M. 1990. Throwing deep: Trade remedy laws in a first-best world. In *Fair exchange: Reforming the trade remedy laws*, edited by M. Trebilcock and R. York. Toronto: CD Howe Institute, pp. 249–50.

U.S. General Accounting Office. 1990. *Internal trade: The use of the GATT Antidumping Code*. Washington, DC: U.S. General Accounting Office.

Viner, J. 1923. *Dumping: A problem in international trade*. Chicago: University of Chicago Press.

26

Strategic Management Journal, Vol. 19, 25–37 (1998)

WHARTON REPROGRAPHICS

CREATING BARRIERS FOR FOREIGN COMPETITORS: A STUDY OF THE IMPACT OF ANTI-DUMPING ACTIONS ON THE PERFORMANCE OF U.S. FIRMS

SARAH J. MARSH*
College of Business, Northern Illinois University, DeKalb, Illinois, U.S.A.

LIPPINCOTT LIBRARY
DOCUMENT DELIVERY

This study investigates whether anti-dumping statutes are effective at improving the performance of U.S. firms. As international trade grows and competitors increasingly cross national borders to enter new markets, U.S. trade law becomes a potentially important tool for managers as they consider how to create barriers for foreign competitors. The results of this study suggest that the anti-dumping laws significantly increase returns of U.S. firms that pursue anti-dumping protection. The average petitioner between 1980 and 1992 received a $46 million increase in market value as a result of filing an anti-dumping petition. However, no significant change in market value was associated with preliminary or final determinations of the International Trade Commission, except when petitions received a negative determination at the final stage of the process. A negative determination at the final stage of the process resulted in a loss of market value. © 1998 John Wiley & Sons, Ltd.

Strat. Mgmt. J., Vol. 19, 25–37 (1998)

> Nowadays, many businessmen, while proclaiming their belief in free trade as a theoretical ideal, tell you at the same time that, as a matter of practical fact, free trade is a myth. Everybody is cheating.
>
> (*The Economist*, 1990: 6)

Creating barriers to competition plays an important role in strategic thinking and practice (D'Aveni, 1994; Porter, 1985). Much of competitive strategy focuses on how to create and sustain barriers that make it difficult for competitors to succeed. A strategy that promotes the development of proprietary knowledge or brand loyalty, or controls sources of supply, for example, focuses on the market factors that create barriers and the potential to generate economic profits (Boddewyn and Brewer, 1994). Corporate political activity can also create effective barriers for competitors (McEvily, Sutcliffe, and Marcus, 1994; Boddewyn and Brewer, 1994). The pursuit of trade protection represents one type of political activity that firms employ to create barriers for foreign competitors and improve the firm's performance.

Many scholars have noted that as tariffs are reduced through multilateral agreements (such as the General Agreement on Tariffs and Trade (GATT)), other types of trade barriers become the preferred mode of protecting domestic firms from foreign competition (Prusa, 1991; Lipson, 1983). This study investigates whether one type of trade barrier, anti-dumping protection, is effective at improving U.S. firm performance. The empirical evidence in this study suggests that the anti-dumping laws significantly increase the returns to U.S. firms that pursue anti-dumping protection.

The following sections of this paper will provide a review of previous studies of trade protection. A brief background on the anti-dumping statutes and the review process will also be pro-

---

Key words: foreign competition; political strategy; trade protection

*Correspondence to: Sarah J. Marsh, College of Business, Northern Illinois University, DeKalb, IL 60115-2854, U.S.A.

CCC 0143–2095/98/010025–13 $17.50
© 1998 John Wiley & Sons, Ltd.

Received 27 March 1995; Revised 17 June 1996; Revised 4 October 1996; Final revision received 7 February 1997

STRATEGIC MANAGEMENT JOURNAL

vided. Then, an overview of the methodology employed in this study, the results, and a discussion will be presented. A detailed description of the methodology can be found in the Appendix.

BACKGROUND

Background literature

Researchers have examined the factors that influence whether firms seek trade protection from foreign competitors and the benefits of pursuing trade protection. Lenway and Rehbein's (1991) study found that organizational slack is a major determinant of involvement in escape clause petitions.[1] Lenway and Schuler's (1991) study of the steel industry tested whether the firm's size, profitability, level of diversification, and competitive strategy affect its investment in trade protection. They found that higher market share helps to explain higher levels of political activity, but they provided only tentative support for the hypothesis that profitable firms are more likely to pursue trade protection. Schuler (1996) also found that high market share firms in the steel industry are more likely to pursue political activity, but he found no relationship between the level of diversification, organizational slack, or the presence of a corporate office in Washington, DC and political activity seeking trade protection. In Salorio's (1991) study, he found that preferences for protection vary based on differences in industry structure, international position, and marketing strategy and channels. These findings are consistent with the view that differences in how firms create value can lead to fundamentally different positions on regulatory issues (Marcus, 1991; Porter, 1985).

The anti-dumping statutes, in particular, have been used extensively by U.S. firms to create barriers for foreign competitors (Boltuck and Litan, 1991). Our understanding of how particular environmental and organizational factors influence the number of anti-dumping petitions is limited, however (Schuler, 1996). Lenway and Schuler's work (1991) suggested that both economic and political factors influence firms' decisions to seek trade protection. Dramatic shifts in the global demand for steel and high levels of import penetration, for example, played a significant role in inducing U.S. steel companies to initiate anti-dumping petitions in early 1982. The outcome of these petitions, however, was affected by political factors that led to negotiated agreements with European Community producers and the withdrawal of the petitions (Lenway and Schuler, 1991). These results conform to Prusa's conclusions that firms may gain protection without pursuing the petition to completion (Prusa, 1991). That is, firms often file anti-dumping petitions and later withdraw them, having achieved either private or government-mediated price and/or quantity agreements with the foreign firms. Thus, strategies to pursue benefits from political activity and competitive strategies may be interdependent (Baron, 1995; Murtha and Lenway, 1994; Ring, Lenway, and Govekar, 1990).

The benefits of pursuing trade protection have been assessed in several studies. In Lenway, Rehbein, and Starks' (1990) study of the steel industry, they found only limited evidence of firm-level wealth effects associated with six trade restrictions on steel import occurring between 1969 and 1982. Lenway and Schuler (1991) found no consistent relationship between the level of involvement in seeking trade protection and its benefits. Lenway, Jacobson, and Goldstein (1990) studied whether firms benefit more from stronger petitions or from lobbying Congress for indirect influence on the International Trade Commission (ITC) decisions; they conclude that corporate resources are better spent improving the strength of the petition rather than lobbying Congress. Thus, while researchers offer theoretical rationale supporting the benefits of seeking trade protection, the evidence quantifying these benefits has been elusive.

This study offers a quantitative evaluation of the effects of pursuing anti-dumping protection. By looking at several stages in the anti-dumping review process, this study attempts to measure the effects of the anti-dumping petition at several stages of the process. By using the firm level of analysis, this study offers an evaluation of impact relevant to managers.

---

[1] Petitions for trade protection through the escape clause seek temporary protection from foreign imports that have injured or threaten to injure a domestic industry. If the International Trade Commission finds injury, a recommendation for import relief is sent to the president, who decides whether to accept, modify, or reject the recommendation (Lenway and Rehbein, 1991).

© 1998 John Wiley & Sons, Ltd.

The review process

The U.S. Department of Commerce (DOC) and the U.S. ITC have broad powers to investigate how U.S. foreign trade influences U.S. production, employment, and consumption (U.S. International Trade Commission, 1992); it is in this role that the DOC and the ITC investigate unfair trade practices, such as dumping, or 'sales at less than fair market value', by foreign firms. 'Sales at less than fair market value' are found when prices charged in the United States by a foreign firm are below the foreign market value. The foreign market value is currently defined in several ways, including the price charged in a foreign firm's home market, prices charged on exports to a third country, or as the total average cost of production (Murray, 1991).

The review process of an anti-dumping petition can take as long as 10–14 months to complete and can result in costly legal bills. (See Figure 1 for a flow chart of the process and the time periods associated with each step.) Petitions are filed by a company or group of companies on behalf of the domestic industry. If the DOC finds a credible allegation on the part of the petitioner(s), the ITC performs a preliminary investigation, determining whether there is a 'reasonable indication' of injury to the industry. If the ITC makes an affirmative ruling of injury, an anti-dumping duty is immediately placed on the imported product (based on the DOC's estimated dumping margin), and foreign firms must post bond or make other guarantees that anti-dumping duties will be paid upon assessment. If the DOC finds a 'reasonable likelihood' in its completed investigation that imports are sold at below fair market value, the ITC then performs a final investigation of whether material injury resulted or is threatened by the pricing practices (U.S. International Trade Commission, 1992; Murray, 1991).

If the DOC's final investigation confirms its earlier ruling that imports are being sold in the

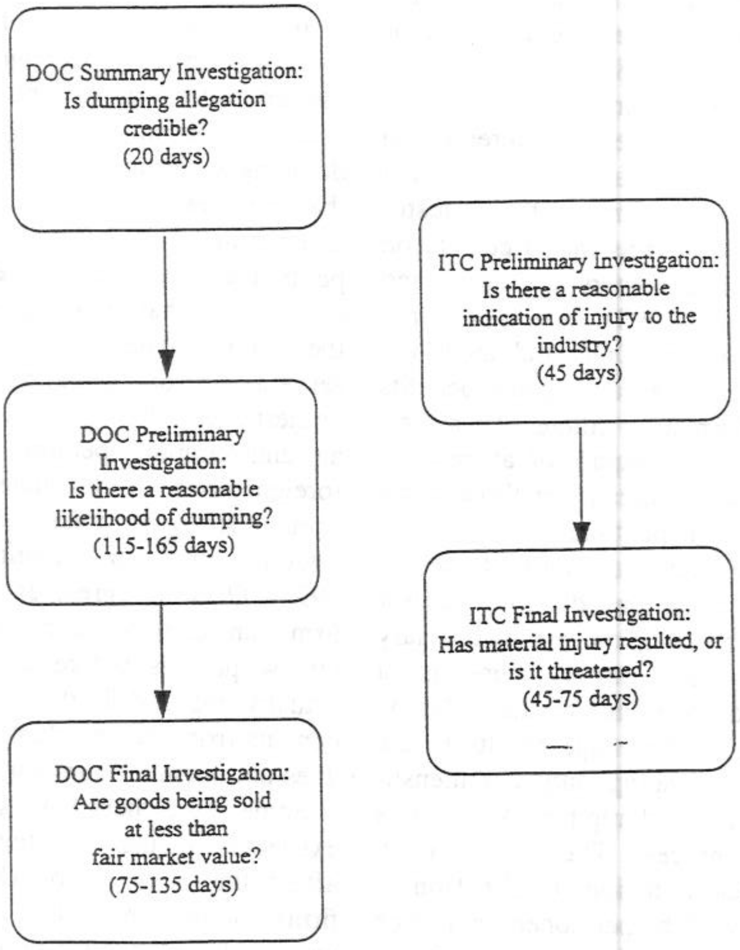

DOC Summary Investigation: Is dumping allegation credible? (20 days)
DOC Preliminary Investigation: Is there a reasonable likelihood of dumping? (115-165 days)
DOC Final Investigation: Are goods being sold at less than fair market value? (75-135 days)
ITC Preliminary Investigation: Is there a reasonable indication of injury to the industry? (45 days)
ITC Final Investigation: Has material injury resulted, or is it threatened? (45-75 days)

Figure 1. Anti-dumping petition review process

© 1998 John Wiley & Sons, Ltd.

United States at less than fair market value, and the ITC's final evaluation finds threatened or actual material injury, then a long-term anti-dumping duty is placed on the imports based on the DOC's estimated margin of dumping. These duties can be imposed unilaterally and uniformly on all product imports from a particular country, or different duty levels can be imposed on each foreign firm. Administrative reviews to assess whether the dumping has continued can be performed annually for up to 3 years following the findings. Petitioners have the right to appeal the decision to the Court of International Trade, which can overturn the decision or require the ITC commissioners to reconsider the petition. Petitioners are also free to continue filing petitions.

RESEARCH HYPOTHESES

All the firms in an industry are potentially affected by the imposition of the anti-dumping duty, but the effects on individual firms may differ due to each firm's strategy, competitive position, or sourcing practices (Porter, 1985; Salorio, 1991). Differences in how goods are produced, how inputs are sourced, and the level of differentiation among the firms' products can create different levels of benefit for each firm in the industry (Salorio, 1991). Although firms may free ride on the political activities of others (Lenway and Rehbein, 1991; Hansen, 1990), one would not expect firms to petition the DOC and incur the costs of the filing unless they anticipated benefits resulting from an affirmative ruling. Therefore, one would expect that the impact of affirmative rulings would be most significant for those firms that file the anti-dumping petitions.

As described in the previous section, there are two different stages in the anti-dumping review process. The ITC's first decision, the preliminary determination, evaluates whether there is a 'reasonable likelihood' of injury to the industry. An affirmative ruling at this stage results in the imposition of an anti-dumping duty commensurate with the amount of dumping. A negative ruling terminates the process. The ITC's second decision, the final determination, results from a more extensive review of the petitioners' evidence of sales at less than fair value. An affirmative ruling at the final stage results in a long-term anti-dumping duty. A negative ruling leads to an elimination of the anti-dumping duty that was imposed after the preliminary ruling, and the process is terminated. Thus, each stage of the ITC review process offers another review of the evidence, additional information about the ITC's evaluation of the evidence, and potentially different levels of anti-dumping protection. Therefore:

> *Hypothesis 1: Affirmative anti-dumping decisions by the ITC at the preliminary and final determinations have positive effects on the performance of the petitioning firms.*

And, conversely:

> *Hypothesis 2: Negative anti-dumping decisions by the ITC at the preliminary and final determinations have negative effects on the performance of the petitioning firms.*

Firms may receive direct and indirect benefits by seeking anti-dumping protection. The most direct benefit of anti-dumping protection is the potential to reduce the threat from foreign competitors by creating barriers to foreign competition through the imposition of an anti-dumping duty. Firms may also benefit indirectly when they file anti-dumping petitions, even if they do not complete the entire review process (Prusa, 1991; Lenway and Schuler, 1991). Simply filing an anti-dumping petition may represent a credible threat to foreign competitors that seek increased market share in the United States. Anti-dumping petitions may serve as a strong signaling device to competitors, domestic and foreign, about price levels. Filing an anti-dumping petition may also induce the foreign firms to negotiate with the petitioning firms directly or with the U.S. government to establish price or quantity agreements (Prusa, 1991). Once an agreement has been reached, the firm can then withdraw the petition from the review process before an anti-dumping duty is actually imposed. In this way, the petitioning firm benefits from the anti-dumping law even though it does not pursue the review process to completion.

Some researchers have suggested that even the existence of the anti-dumping law itself may affect the behavior of U.S. firms and foreign firms (Stern, 1991). Just the *threat* of imposing an anti-dumping duty may reduce price competition by foreign competitors and induce them to

© 1998 John Wiley & Sons, Ltd.

raise prices to customers, making the domestic firms' products more price competitive.

Anti-dumping actions may also be sought to gain competitive advantage over *domestic* competitors. Different industry structures can make cooperative pricing more or less difficult, and technological changes, supply disruptions, and other exogenous factors can affect the relative position of competitors. Significant intraindustry conflict can occur because the potential effects of trade protection may differ for the individual firms in the industry (Salorio, 1991), thereby creating the potential to use trade protection as a means to establish competitive advantage *vis-à-vis* domestic competitors.

Thus, there may be direct and indirect benefits associated with pursuing anti-dumping protection. Firms may receive some of these benefits, such as increased power in private and/or government-led negotiations and signaling of prices to foreign or domestic competitors, even if the petition does not proceed through the entire process successfully. Therefore:

> *Hypothesis 3: Filing an anti-dumping petition has a positive effect on the performance of petitioning firms, regardless of the outcome of the petition.*

METHODOLOGY

Several considerations led to the choice of an event study methodology for this research. First, the public announcements of petition filings and the ITC's decisions offer clear and definitive events to which the market's response can be measured. The preliminary and final determination decisions made by the ITC are publicly announced upon decision. Second, market measures of performance offer a direct measure of shareholder value—one which may more readily represent the value of the firm than traditional accounting measures of performance (Lubatkin and Shrieves, 1986). In this case, accounting returns may not necessarily capture the value associated with using the anti-dumping law as part of a firm's competitive strategy. For example, an anti-dumping action may help sustain price levels at their current level or induce foreign competitors to enter negotations on prices and quantities, thereby creating value for the firm that would otherwise not be created—value that accounting measures would not reliably capture.

According to the efficient markets hypothesis, market prices represent the net present value of all future cash flow and reflect all available information about the value of a security; a semi-strong form of efficiency assumes that new information is quickly assimilated and reflected in the security price (Brealey and Myers, 1991). In this study, like other event studies (e.g., Bettis and Weeks, 1987; Lubatkin, 1987), any change in price that cannot be accounted for in general market moves or firm risk is assumed to reflect a change in anticipated cash flow in the future that is associated with new information presented by a particular event.

This research employed an event study methodology to isolate the unanticipated change or abnormal return associated with anti-dumping petitions. Three different events were tested: the date the petition was initially filed, the date of the preliminary ITC determination, and the date of the final ITC determination. A 3-day event period, extending from 2 days prior to the event until the day of the event, was analyzed. To test the robustness of the 3-day event period results, additional analysis using a 5-day event period, extending from 3 days prior to the event until 1 day following the event, was also performed. Each of these event periods allows for the market response to leakage of information to occur before the event. See the Appendix for a detailed description of the methodology.

Sample

Between 1980 and 1992, 626 petitions were filed with the Department of Commerce alleging dumping practices by foreign firms. The petitioners during this period consisted of over 336 different organizations, including privately held and publicly traded companies, trade associations, *ad hoc* committees and industry groups, and labor unions.

The sample used in this study was constructed by identifying the petitioners of the 626 filings reported in the *USITC Annual Reports* for fiscal years 1980 through 1992. Fiscal year 1980 represents the first year of implementation of the Trade Agreement Act of 1979 which made significant changes to U.S. anti-dumping law (Prusa, 1991), and 1992 was the most recent data avail-

© 1998 John Wiley & Sons, Ltd.

able at the commencement of this study. The petitioners were identified by cross-referencing the ITC petitions with those recorded by the International Trade Administration of the Department of Commerce. Petitioner names were then retrieved from the Federal Register notices announcing the initiation of the anti-dumping investigation. Data on the trading status of the petitioners, including parent company names, were then collected using *Directory of Corporate Affiliations* and Dun and Bradstreet's *America's Corporate Families*. Dates for the three events tested in this study—the initial filing of the petitions, preliminary determinations, and final determinations—were retrieved from the *U.S. ITC Annual Reports*. Companies that experienced potential confounding events, including surprise earnings announcements and engagement in hostile takeover bids, during the event period were identified using the *Wall Street Journal Index* and eliminated from the sample.

Publicly held companies (or their majority-owned subsidiaries) that trade on the New York, American, or NASDAQ stock exchanges accounted for 147 filing events over the study period. Every publicly held company that petitioned the ITC was included. After eliminating nine filing events for missing financial return data and four filing events for confounding events, 134 filing events by 88 publicly held companies were in the final sample. Because firms file multiple petitions simultaneously, the 134 filing events in the final sample represent 51 percent (316) of the total universe of petitions. Analysis of the preliminary and final determination events included 133 petitions that completed the preliminary determination stage and 72 petitions that reached the final determination stage. Analysis of the petitions based on the ITC decisions (i.e., affirmative decisions or 'winners' and negative decisions or 'losers') was performed for the initial filing, preliminary determination, and final determination events.

Table 1 provides a partial listing of the petitions, products, and petitioners in the sample. The average market capitalization of petitioning firms was \$3.0 billion and ranged from \$5 million to \$33 billion. The median capitalization was \$980 million.

RESULTS

Table 2 presents the summarized results of the analysis using a 3-day event period. The results show that for the initial filing of an anti-dumping petition, the average abnormal return was 0.01, or 1 percent, which is statistically significant ($p < 0.01$). The results associated with preliminary and final determinations, on the other hand, show small and insignificant impacts on the market price of the security of 0.1 and –0.5 percent, respectively.

Analysis of the change in market valuation was performed for the initial filing event. Using the current shares outstanding, the closing security price on the day prior to the event window, and the change in security price over the event window, the average of the changes in market capitalization was calculated. The analysis suggests that an initial filing event results in an average increase of \$45.9 million in market capitalization for petitioning firms.[2]

Affirmative findings at both preliminary and final determinations show small and insignificant impacts (0.1% and 0.2%) on the market price of the security. These results do not support the first hypothesis that predicted that affirmative decisions by the ITC would have positive effects on the petitioning firms.

The results also show that negative findings by the ITC at the final determination produced significant abnormal returns, supporting the second hypothesis that predicted that negative anti-dumping decisions by the ITC would have negative performance effects on petitioning firms. Table 2 shows that firms with petitions that received negative final determinations ('losers') had a –2.1 percent change in market return ($p < 0.01$). The strength of this finding is enhanced by the relatively low power of the test resulting from the small number of observations (20) in this subsample. However, firms that received a negative preliminary determination had

---

[2] Note that based on the average abnormal return of 0.01 and the average market capitalization of \$3 billion, the average firm in this sample would anticipate a \$30 million change in market capitalization. The difference between the change in market capitalization for an average firm and the average change in this sample reflects the fact that the product of the means (of the changes in security prices and in market capitalizations) does not equal the mean of the products (of each firm's change in security price and market capitalization).

© 1998 John Wiley & Sons, Ltd.

Table 1. Partial listing of petitions, products, and petitioners

Petition no.	Product	Petitioner(s)[a]
731-TA-31	Barium carbonate and strontium carbonate	FMC Corp.
731-TA-38	Truck trailer axle-and-brake assemblies	Rockwell International Corp.
731-TA-53	Hot-rolled carbon steel plate	Cyclops Corp.
731-TA-87	Seamless steel pipes and tubes	Babcock & Wilcox Company
731-TA-94	Bicycle tires and tubes	Carlisle Tire and Rubber Co.
731-TA-107	Melamine	Melamine Chemicals, Inc.
731-TA-108	Portland cement	Kaiser Cement Corp. Gifford-Hill Cement Co.
731-TA-119	Lightweight polyester filament fabric	J. P. Stevens & Co., Inc.
731-TA-165	Valves, nozzles, and connectors of brass	Badger-Powhatan, Inc.
731-TA-200	Radial-ply tires for passenger cars	Armstrong Rubber Co.
731-TA-207	Cellular mobile telephones and subassemblies	Motorola, Inc.
731-TA-238	12-volt motorcycle batteries	General Battery Corp.
731-TA-240	Photo albums and filler pages	Kleer-Vu Plastics Corp.
731-TA-288	Erasable programmable read-only memories	Intel Advanced Micro Devices National Semiconductor Corp.
731-TA-297	Porcelain-on-steel cooking ware	General Housewares Corp.
731-TA-341	Tapered roller bearings, parts, and housings	Timken Co.
731-TA-355	Silica filament fabric	Ametek, Inc.
731-TA-486	Coated groundwood paper	Bowater Inc.
731-TA-522	Minivans	Ford Motor Co. Chrysler General Motors Corp.
731-TA-623	Hairbrushes, parts, and components	Goody Products, Inc.

[a]Petitioners may include privately held and publicly traded companies, trade associations, *ad hoc* committees and industry groups, and labor unions. Only publicly traded companies that were petitioners in the sample are shown in the partial listing of petitions listed here.

Table 2. Test results

Type of event and subsample	Average abnormal return	*Z*-Statistic	Sample size[a]
Initial filing of petition	0.010	2.57**	134
Eventual winners in preliminary determination	0.011	2.59**	115
Eventual losers in preliminary determination	0.002	0.20	17
Eventual winners in final determination	0.011	1.25	48
Eventual losers in final determination	0.011	1.57†	19
Preliminary determination	0.001	0.17	133
Winners	0.001	0.18	115
Losers	0.004	0.93	18
Final determination	–0.005	–0.75	72
Winners	0.002	0.64	52
Losers	–0.021	–2.45**	20

†$p < 0.10$; **$p < 0.01$

[a]Changes in the size of the samples result from termination of petitions as a result of ITC decisions, withdrawal of petitions by petitioners, and missing financial and ITC data.

© 1998 John Wiley & Sons, Ltd.

a small and insignificant (0.4%) increase in market return—a result that fails to support the second hypothesis.

The support for the third hypothesis is mixed. When a firm files a petition, the average effect is 0.01, or 1 percent, ($p < 0.01$). Positive abnormal returns of 0.01 ($p < 0.01$) were found for those petitions that were eventual winners in the preliminary determination. Even the eventual losers in the final determination stage showed positive and marginally significant abnormal returns of 0.011 ($p < 0.10$). In addition, those firms that received a negative determination at the preliminary stage did not suffer a strong negative effect on performance. Each of these findings supports the hypothesis that firms benefit from filing anti-dumping petitions, regardless of the petition's outcome. On the other hand, those firms that receive a negative final determination suffered a strong negative effect—one which was nearly twice as large as the positive effect of the initial filing. Thus, the net wealth effect is contingent on the eventual petition outcome, but the market does not distinguish eventual winners and losers at the time of initial filing.

The steel industry has consistently pursued and received administrative trade protection (Prusa, 1991; Boltuck and Litan, 1991). Because the repeated pursuit of trade protection leads potentially to a better understanding of the review process and its potential consequences for the industry and individual firms by the petitioners and the financial markets, subsample analysis was performed to assure that the steel industry did not dominate the results. Subsample analysis of the petition events reveals that of the 134 events associated with the initial petitions, 52 were for steel products (not shown). For the steel subsample, the average cumulative abnormal return over the event period was 1.1 percent ($p < 0.10$). For the 82 petitions associated with nonsteel products, the average abnormal return was 0.9 percent ($p < 0.05$). Standard t-tests show no significant difference between the abnormal returns of the two groups, however.

The analysis using the 5-day period supported the results using the 3-day event period.

DISCUSSION

This analysis suggests that the market anticipates that anti-dumping petitions will create increased cash flow, and as a result, the market price of the securities increases 1 percent beyond the expected normal market price. The positive effect of anti-dumping actions occurred at the time of the initial filing, before the preliminary and final determinations were made. Even those firms that eventually received a negative determination at the final stage of the process experienced a positive effect when the petition was initially filed. This positive effect is not surprising, given the high proportion of successful petitions. Over the 13-year period, 85 percent of the sample received affirmative findings in the ITC preliminary determination, resulting in the imposition of anti-dumping duties on foreign competitors. The high proportion of successful petitions may account for the lack of consistently significant returns associated with the preliminary and final determinations. In essence, due to the high success rate of anti-dumping petitions, the market responds to the filing, not to the determinations themselves.

On the other hand, no significant increases in security prices resulted from affirmative findings at either the preliminary or final determinations. In addition, no significant change resulted from a negative preliminary determination. However, firms that successfully pursued cases through the final stage, but then received a negative final determination, suffered strong and significant negative effects on the security price. A negative finding produces an average negative effect of 2.1 percent compared to the average 1.1 percent increase that was experienced by these same firms when the petition was filed.

Thus, these results suggest that ITC decisions influence security prices, although the influence of these decisions is not consistent across all stages of the process. Some of the benefits of pursuing trade protection may not depend on ITC decisions. These benefits may be associated with signaling prices or bringing foreign competitors to the negotiating table, for example.

One explanation for the strong negative effect of negative final determinations may result from how firms proceed through the anti-dumping review process. For example, once the initial filing has occurred, some firms will enter negotiations with foreign competitors to resolve the petition without proceeding through the entire review process. If the petition can be resolved, the petitioning firm(s) will then terminate the process (Prusa, 1991; Lenway and Schuler, 1991).

© 1998 John Wiley & Sons, Ltd.

Thus, the cases that proceed onto the final stage of the process represent cases where petitioning and foreign firms could not come to a mediated agreement. This process of 'selecting out' may contribute to the strong negative effect of an unfavorable final determination.

A simplistic interpretation of these results would suggest that companies should 'file early and file often', but such advice would be premature. First, this study's results represent an average effect of all petitions filed over a 13-year period; and, while the length of the study period and wide range of firms in the study increases the generalizability of the *average* results, the application of these results to individual firms may prove unwise. A wide variety of industries, ranging from steel products to hairbrushes, are represented in the sample (see Table 1). While the steel and nonsteel subsamples were analyzed separately, further investigation would be necessary to understand how firm size, industry structure, and other competitive factors may influence the outcome associated with an anti-dumping petition.

Second, the benefits of pursuing trade protection found in this study are based on a sample that includes only the publicly traded firms that initiated petitions on their industry's behalf; those firms that filed petitions as part of an industry group or trade association are not represented. By testing events associated with only those firms that directly petitioned the ITC (and not those who petitioned as members of an industry group or trade association), this test focuses on firms that are likely to enjoy the greatest benefits from a successful petition. Large and high market share firms may receive more benefits from trade protection, because they have more at risk if they fail to deter foreign competition (Lenway and Schuler, 1991; Schuler, 1996); and, therefore, they may be more likely to pursue trade protection individually. Smaller firms may have less to gain from trade protection than larger firms, and they may have lower expectations about their ability to successfully acquire trade protection. In addition, applying the average 1 percent increase in market valuation to a smaller firm (say with a valuation of $50 million) would create a benefit of only $500,000. Discussions with previous participants by this researcher and others (Schuler, 1996) suggest that this smaller benefit may be well below the legal costs associated with pursuing a petition through the administrative process. Smaller firms may try to maximize the net benefits of trade protection by minimizing its costs by petitioning through trade associations or *ad hoc* industry groups. The level of firm-level benefit associated with filing anti-dumping petitions found in this research may not apply to smaller firms.

This study also does not answer whether or how much the nonparticipating firms benefited from the anti-dumping petitions pursued by other firms in the industry. Some firms may free ride on the political activity of others, and they may benefit nonetheless if trade protection is granted. Other nonparticipating firms may be indifferent to or oppose trade protection, based on expectations about the overall effect of trade protection (Salorio, 1991; Schuler, 1996). Thus, conclusions about the benefits of trade protection for nonparticipants are difficult to make on the basis of this study.

Finally, the loss in market value that occurs when a petition receives a negative ruling at the final stage of the process is large and twice the magnitude of the positive effect of the initial filing. Unless the petition is credible, the firm may suffer considerable loss in market value. These potential costs should not be ignored by managers.

Further investigation of specific industry and firm effects would increase our understanding of the benefits of anti-dumping protection and how they are distributed among firms in an industry. Differences in levels of industry concentration, vertical linkages, product differentiation, and other factors may impact the potential benefits of anti-dumping protection. Differences in these competitive factors can create rivalries in political activities, just as they influence patterns of economic competition (Salorio, 1991).

By investigating how industry and firm-level differences affect the outcomes associated with anti-dumping petitions, we can also gain insight into how these outcomes may compare with the other strategic options that firms pursue to improve their competitive position in the U.S. market. Firms that have high exit costs and/or high downsizing costs, for example, may benefit more from trade protection than other firms, and, therefore, may pursue it more often (Schuler, 1996). For these firms, trade protection may produce a larger net benefit when compared with other strategies that create barriers for foreign competitors. Firms with lower exit costs and/or lower downsizing costs may pursue options other than trade protection (such as moving production offshore), to improve their competitive position. Firms in industries that are concentrated may also be more likely to pursue trade protection because their smaller numbers make it easier and less costly to organize their efforts toward a common objective (Olson, 1965). In addition, because the petitioner(s) must demonstrate to the ITC material injury to the industry, petitions may be strengthened when a larger proportion of the industry participates in petitions for trade protection. Thus, different industry structures and strategic capabilities may expand or contract the firm's range of potential responses to increased competition by foreign competitors and influence when anti-dumping protection is sought and its ultimate effect.

Future studies should also employ alternative methods to measure firm-level benefits. Although the event study methodology employed in this study is well accepted as a means of valuing changes in future cash flows, the capital asset pricing model on which it is based assumes the market's ability to accurately assess the impact of new information on the firm. In addition, other methodological approaches (using interviews and/or surveys, for example), may also help identify the sources of value that contribute to the increased financial returns found in this study. Firms may benefit from anti-dumping protection due to an increased ability to signal or increase prices, or due to increased market share, or a combination of factors.

In-depth qualitative research may shed light on how experience with the process affects managers' use of U.S. trade law and its effects. For example, there is some evidence that successful petitions may have negative consequences for U.S. firms that are trying to enter or expand in the foreign firm's home market; that is, the imposition of anti-dumping duties may attract the attention of officials in foreign governments and may elicit retaliatory actions that make it more difficult for U.S. firms to succeed in foreign markets (Awanohara, 1993). Political activity may also be perceived negatively in the United States, potentially damaging the reputation of the firm (Schuler, 1996). Other methodological approaches to the topic may improve our understanding of the more indirect consequences of seeking administered trade protection.

This study measures the effects of anti-dumping actions at the firm level. Consequently, conclusions about public policy cannot be made directly. This research, however, does point to several areas worthy of consideration by policy-makers. First, the results suggest that firms benefit significantly from anti-dumping petitions; but, it is unclear whether the overall welfare benefits to firms, their shareholders, their employees, their suppliers, or other stakeholders are justified in terms of the potentially higher costs to U.S. consumers. Second, the size of the change in market capitalization, over $46 million, suggests that significant benefits accrue to shareholders as a result of anti-dumping petitions. These findings suggest that some benefit goes directly to the firms' shareholders in the form of higher stock prices, contradicting scholars who have suggested that the benefits of anti-dumping protection go only to the government in the form of duty collections and that the process yields no benefit for injured firms (Tyson, 1993). Finally, the results of this study raise questions about the anti-dumping review process itself. For instance, firms may use the anti-dumping process to pursue indirect benefits, such as mediated agreements. It is unclear whether using this process to achieve these benefits is the most efficient means of resolving conflicts between domestic and foreign firms.

As international trade grows, the need to improve our understanding of how firms can compete effectively with foreign competitors also increases. More research is needed to understand what types of activity improve performance and how political activity complements other strategic decisions and actions a firm makes. Previous research has offered theoretical rationale supporting the benefits of seeking trade protection, but the empirical evidence quantifying these benefits has been limited. This study offers empirical evidence that firms can create barriers for foreign competitors and benefit significantly by seeking trade protection.

ACKNOWLEDGEMENTS

I would like to thank Rich Bettis, Hugh O'Neill, Mike Lord, Adelaide Wilcox King, Sally Fowler,

© 1998 John Wiley & Sons, Ltd.

Chris Shook, and two anonymous referees for their helpful comments on earlier drafts of this paper. I would also like to acknowledge financial support for this research provided by the International Business Education Center, which is jointly administered by the Kenan-Flagler Business School of the University of North Carolina at Chapel Hill and the Frank Hawkins Kenan Institute of Private Enterprise.

REFERENCES

America's Corporate Families (1980–1992 annual volumes). Dun and Bradstreet, Parsipanny, NJ.

Awanohara, S. (15 September, 1993). 'Currency risk: Japan's firms raise U.S. prices to avoid dumping charges', *Far Eastern Economic Review*, pp. 75–76.

Baron, D. P. (1995). 'Integrated strategy: Market and nonmarket components', *California Management Review*, **37**(2), pp. 47–65.

Bettis, R. A. and D. Weeks (1987). 'Financial returns and strategic interaction: The case of instant photography', *Strategic Management Journal*, **8**(6), pp. 549–563.

Boddewyn, J. J. and T. L. Brewer (1994). 'International-business political behavior: New theoretical directions', *Academy of Management Review*, **19**(1), pp. 119–143.

Boltuck, R. and R. E. Litan (1991). 'America's "unfair" trade laws.' In R. Boltuck and R. Litan (eds.), *Down in the Dumps: The Administration of Unfair Trade Laws*. Brookings Institution, Washington, DC, pp. 1–22.

Brealey, R. A. and S. C. Myers (1991). *Principles of Corporate Finance* (4th ed.). McGraw-Hill, New York.

Brown, S. J. and J. B. Warner (1985). 'Using daily stock returns: The case of event studies', *Journal of Financial Economics*, **14**, pp. 3–31.

D'Aveni, R. A. (1994). *Hyper-Competition: Managing the Dynamics of Strategic Maneuvering*. Free Press, New York.

Directory of Corporate Affiliations (1980–1992 annual volumes). National Register Publishing Co., Wilmette, IL.

Economist (22 September, 1990). 'Jousting for advantage' (in Survey of World Trade), pp. 5–6.

Hansen, W. (1990). 'The international trade commission and the politics of protectionism', *American Political Science Review*, **84**(1), pp. 21–46.

Lenway, S. A., C. K. Jacobson and J. Goldstein (1990). 'To lobby or to petition: The political environment of U.S. trade policy', *Journal of Management*, **16**(1), pp. 119–134.

Lenway, S. A. and K. Rehbein (1991). 'Leaders, followers, and free riders: An empirical test of variation in corporate political involvement', *Academy of Management Journal*, **34**(4), pp. 893–905.

Lenway, S. A., K. Rehbein and L. Starks (1990). 'The impact of protectionism on firm wealth: The experience of the steel industry', *Southern Economic Journal*, **56**(4), pp. 1079–1093.

Lenway, S. A. and D. A. Schuler (1991). 'Corporate political involvement in trade protection'. In R. E. Baldwin (ed.), *Empirical Studies of Commercial Policy*. University of Chicago Press, Chicago, IL, pp. 75–105.

Lipson, C. (1983). 'The transformation of trade: The sources and effects of regime change'. In S. D. Krasner (ed.), *International Regimes*. Cornell University Press, Ithaca, NY, pp. 233–271.

Lubatkin, M. (1987). 'Merger strategies and stockholder value', *Strategic Management Journal*, **8**(1), pp. 39–53.

Lubatkin, M. and R. E. Shrieves (1986). 'Towards reconciliation of market performance measures to strategic management research', *Academy of Management Review*, **11**(3), pp. 497–512.

Marcus, A. A. (1991). 'Airline deregulation, business strategy, and regulatory theory'. In M. J. Dubnich and A. R. Gitelson (eds.), *Public Policy and Economic Institutions*. JAI Press, Greenwich, CT, pp. 325–350.

McEvily, S., K. Sutcliffe and A. Marcus (1994). 'Explaining preferences for public policy: Strategic and contextual determinants', *Academy of Management Best Paper Proceedings*, pp. 319–323.

Mikkelson, W. H. and M. M. Partch (1986) 'Valuation effects of security offerings and the issuance process', *Journal of Financial Economics*, **15**, pp. 31–60.

Murray, T. (1991). 'The administration of the Antidumping Duty Law by the Department of Commerce.' In R. Boltuck and R. Litan (eds.), *Down in the Dumps: The Administration of Unfair Trade Laws*. Brookings Institution, Washington, DC, pp. 23–63.

Murtha, T. P. and S. A. Lenway (1994). 'Country capabilities and the strategic state: How national political institutions affect multinational corporations' strategies', *Strategic Management Journal*, Summer Special Issue, **15**, pp. 113–129.

Olson, M. (1965). *The Logic of Collective Action: Public Goods and the Theory of Groups*. Harvard University Press, Cambridge, MA.

Porter, M. E. (1985). *Competitive Advantage*. Free Press, New York.

Prusa, T. J. (1991). 'The selection of antidumping cases for ITC determination.' In R. E. Baldwin (ed.), *Empirical Studies of Commercial Policy*. University of Chicago Press, Chicago, IL, pp. 47–71.

Ring, P. S., S. A. Lenway and M. Govekar (1990). 'Management of the political imperative in international business', *Strategic Management Journal*, **11**(2), pp. 141–151.

Salorio, E. M. (1991). 'Trade barriers and corporate strategies: Why some firms oppose import protection for their own industry', unpublished doctoral dissertation, Harvard University.

Schuler, D. A. (1996). 'Corporate political strategy and foreign competition: The case of the steel industry', *Academy of Management Journal*, **39**(3), pp. 720–737.

© 1998 John Wiley & Sons, Ltd.

Stern, R. M. (1991). 'Comment on "The selection of antidumping cases for ITC determination"'. In R. E. Baldwin (ed.), *Empirical Studies of Commercial Policy*. University of Chicago Press, Chicago, IL, pp. 71–74.

Tyson, L. D. (1993). 'Who's bashing whom? Trade conflict in high-technology industries', Institute for International Economics, Washington, DC.

U.S. International Trade Commission (1980–1992 annual volumes). *United States International Trade Commission Annual Report*. Government Printing Office, Washington, DC.

APPENDIX

According to the efficient markets hypothesis, equity market prices represent the net present value of all future cash flows. Any unanticipated change or abnormal return can be isolated using the market model which assumes a linear relationship between the expected return of an individual stock and the expected return of a broadly-based market portfolio or index:

$$R_{jt} = \alpha_j + \beta_j R_{mt} + \epsilon_{jt} \qquad (1)$$

where: R_{jt} = rate of return of firm j over period t

β_j = systematic risk of security j over period t

R_{mt} = rate of return of a value-weighted market portfolio over period t

ϵ_{jt} = random error of rate of return on security j over period t

Ordinary least squares is used to estimate α, β, and s_ϵ (standard deviation of the error term) under the assumption that the error terms have an expected value of zero and are not correlated. In theory, the error term, ϵ_{jt}, captures any variance that cannot be explained by the overall market, and therefore, the error represents the change in price associated with the event.

In order to estimate the anticipated, or normal return, for a security, the parameters α and β are estimated for a period prior to the event. In this study, the period over which the normal return is estimated is from 250 trading days prior to the event to 50 trading days prior to the event. Estimating the normal return well before the event dates makes it less likely that information about the upcoming event has already been incorporated into the normal return of the security. Estimating the normal return over a period of 200 days or more is consistent with most financial research (e.g., Brown and Warner, 1985), and it provides a meaningful basis for the comparison of abnormal returns associated with a given event.

Given the OLS parameter estimates, $\hat{\alpha}$ and $\hat{\beta}$, the abnormal return for the period t is calculated as:

$$AR_{jt} = R_{jt} - (\hat{\alpha}_j + \hat{\beta}_j R_{mt}) \qquad (2)$$

where R_{jt} is the actual return for a given security and $\hat{\alpha} + \hat{\beta}R_{mt}$ is the market-adjusted, or normal, return for a security over the event period. AR, or the abnormal return, represents the abnormal change in price that cannot be attributed to the general market or firm risk, and therefore, is attributed to the event that occurred. In most event studies, the event period tightly wraps the event under study. By shortening the length of time prior to the event, the opportunity for nonrelated events to confound the results is minimized. A short event period also increases the power of the test (Brown and Warner, 1985).

Tests of statistical significance of the abnormal return are based on standardized abnormal returns; thus, the standardized abnormal return on a given day is defined as (Mikkelson and Partch, 1986):

$$SAR_{jt} = \frac{AR_{jt}}{S_{jt}} \qquad (3)$$

where:

$$S_{jt} = \sqrt{V_j^2\left[1 + \frac{1}{ED} + \frac{(R_{mt} - \overline{R}_m)^2}{\sum_{i=1}^{ED}(R_{mi} - \overline{R}_m)}\right]} \qquad (4)$$

V_j^2 is the residual variance of firm j's market model regression, ED is the number of days in the period used to estimate the market model, R_{mt} is the market return on day t, and R_m is the mean market return in the estimation period. Because the standardized abnormal return is calculated for a given day, a cumulative abnormal return is necessary for an event period that consists of more than 1 day. In addition to the aggregation of the standardized abnormal returns for each firm over the event period, the sample of observations (that consists of multiple firms

© 1998 John Wiley & Sons, Ltd.

with events occurring on multiple dates) is aggregated for all firms in order to determine if the abnormal returns of the sample are statistically different from zero, that is, whether the events created a significant change in security price. Assuming cross-sectional independence, the average standardized cumulative abnormal return (*ASCAR*) for a sample of observations for any time period is (Mikkelson and Partch, 1986):

$$ASCAR_{T_1T_2} = \frac{1/N\sum_{j=1}^{N}\sum_{t=T_1}^{T_{2j}} SAR_{jt}}{\sqrt{T_{2j} - T_{1j} + 1}} \qquad (5)$$

where: N = the number of firms in the sample
T_1 = beginning of event period
T_2 = end of event period
j = firm index

Assuming that the average standardized cumulative abnormal return has variance equal to $1/N$ and is symptotically normally distributed, the Z-statistic can be used to test for significance:

$$Z = \sqrt{N}\,(ASCAR_{T1.T2}) \qquad (6)$$

Finally, the change in the market valuation associated with an event can also be estimated by multiplying the firm's market capitalization on the day preceding the event window by the cumulative standardized abnormal return over the event period.

The source for the security returns was the daily return tape compiled by the Center for Research in Security Prices at the University of Chicago. The market returns were calculated using a value-weighted index of all securities traded on the NYSE, AMEX, and NASDAQ exchanges.

© 1998 John Wiley & Sons, Ltd.

LIPPINCOTT LIBRARY
DOCUMENT DELIVERY SERVICE

WHARTON [illegible]GRAPHICS

COMMENTARY

Brave New World: International Regulation of Subsidies and the Future of State and Local Economic Development Programs

Charles S. Colgan
University of Southern Maine

Recent international trade agreements are likely to have profound effects on state and local development programs in coming years. The most important changes will come as a result of the new Agreement on Subsidies and Countervailing Measures concluded as part of the General Agreement on Tariffs and Trade. Provisions of the North American Free Trade Agreement and the United States-Canada Free Trade Agreement may also affect development efforts. This article examines the new rules on subsidies and their implications for state and local economic development programs. It concludes that these programs are likely to come under increasing attack from trading partners of the United States. While there are limited exceptions to the new disciplines, many programs may have to be changed to comply with the new rules. Some programs may face countermeasures in other countries or may even be required to be removed by international dispute settlement procedures.

Two major international trade policy agreements reached in 1993 are likely to have profound effects on state and local development programs in the future. The most important changes will come as a result of the new Agreement on Subsidies and Countervailing Measures concluded as part of the Uruguay Round of multilateral trade negotiations under the auspices of the General Agreement on Tariffs and Trade (GATT). Provisions of the North American Free Trade Agreement (NAFTA) and its predecessor, the United States-Canada Free Trade Agreement of 1988, may also affect development efforts in various ways. This article examines the new rules on subsidies and their implications for state and local economic development programs in order to provide a brief introduction to the issues that development policymakers and administrators will confront in coming years.[1]

State and local economic development efforts are increasingly focused on international trade and investment as the American economy becomes more integrated into a global economy. Evidence for this trend is abundant. At the beginning of the 1980s, only two states had trade offices abroad; at the end of the decade, nearly 40 states did. At the end of the 1980s, there were more states with offices in Tokyo than in Washington. Governors and mayors of large cities routinely travel abroad on trade missions promoting products of their states or cities and seeking investment from abroad.[2]

The new focus on international trade in economic development will bring state and local economic agencies into contact with a new world of international trade law and policy that has traditionally been the sole concern of the federal government. One of several international trade

Charles S. Colgan is Associate Professor of Public Policy and Management, Edmund S. Muskie Institute of Public Affairs, University of Southern Maine. He was formerly the Maine State Economist, Trade Policy Consultant—Office of the Governor of Maine. He has a B.A. from Colby College and a Ph.D. from the University of Maine.

ECONOMIC DEVELOPMENT QUARTERLY, Vol. 9 No. 2, May 1995 107-118
© 1995 Sage Publications, Inc.

108·

rules with which these governments will have to become familiar has direct potential effects on economic development policies: the discipline of "unfair trade practices" and, in particular, subsidies. Economic development assistance to businesses has the potential to bring those businesses, and the development programs, real problems when they enter foreign markets, and perhaps even when they compete with foreign products at home.

1. THE GATT-1994

The GATT was established in 1947 as an international mechanism for reducing trade barriers and promoting cooperative negotiated solutions to trade disputes. Under its auspices several rounds of negotiations have been conducted to lower tariffs and establish rules governing trade. The latest, known as the Uruguay Round after the country in which the talks were launched in 1986, will result in reduction of tariffs and such a comprehensive overhaul of GATT rules that the entire agreement has been renamed the "GATT-1994." The GATT-1994 will cover more of world trade than ever before, with 118 countries expected to become members.

The question of subsidies' effects on trade has been a focus of GATT negotiations for a number of years. The first negotiations on how to discipline the use of subsidies culminated in the 1979 GATT Subsidies Code, part of the agreements concluded in the Tokyo Round of multilateral trade negotiations held during the 1970s. However, this first attempt was at best a partial success. Agreement was limited to prohibiting export subsidies (those designed solely to encourage exports), since they had the clearest effects on trade. In defining what constituted a trade-distorting "domestic subsidy"—that is, a subsidy to a firm offered for general development purposes and not necessarily tied to any expectation of exporting—the Tokyo Round reached only "straightforward ambivalence."[3] No agreement could be reached on what should distinguish a legitimate government expenditure from a trade-distorting subsidy. The Tokyo Round Code thus could do no better than to discourage the use of domestic subsidies if they affect trade between GATT members.[4] This failure to deal with the domestic-subsidy issue resulted in a decade of continuing disputes between the United States and its major trading partners. Resolving this issue through a clearer definition of domestic subsidies and more certain disciplines on those that distort trade became a major element in the negotiating agenda of the United States in the Uruguay Round.

Other countries also had agendas for the subsidies negotiations. Canada and the European Community (EC) wanted exceptions to any new definitions of a domestic subsidy to explicitly permit certain policies they considered necessary to their economies, including efforts at "regional development." In addition to the issue of definitions, all the major countries wanted agreement on procedures to be used when countering subsidies, at both international and national levels, that would ensure consistency of process and certainty of discipline.

The Agreement on Subsidies and Countervailing Measures in the GATT-1994 largely met everyone's objectives, greatly expanding international rules on subsidies and clarifying what is meant by "subsidy," when a subsidy is to be considered likely to affect international trade, and what remedies may be applied.

Definitions

The GATT-1994 contains a definition of "subsidy" drawn largely from U.S. law and practice. Two elements determine which subsidies are of concern: the *type of government action* and the *specificity* of its application. A very wide range of government activities in support of businesses is covered by the definition, including

- grants,
- loans,
- equity,
- loan guarantees,

109

- forgiveness of taxes otherwise due,
- provision of goods or services other than infrastructure, and
- government purchase of goods.

The critical factor in this element of the definition is whether any "benefit" is conferred by the government on the receiving firms or industries and the extent of that benefit. Measuring the extent of the benefit conferred is a critical factor in determining whether the total subsidies exceed certain thresholds that determine presumptions of the likelihood of trade-distorting effects. These measurements are also important in determining the amount of any duties levied to offset (countervail) the effects of subsidized imports. The agreement provides guidance about how the value of subsidies is to be measured for principal forms of assistance:

The value of *equity* subsidies is taken to be the difference between the government action and the usual investment practice of private investors in the country (adjusted for risk).

Loan subsidies are valued as the difference between payments on the government loan and those on "a comparable private loan which the firm can actually obtain in the market."[5]

Loan guarantee value is the difference between the payment for a guaranteed loan and what would be paid for a comparable private loan absent the guarantee.

If the *government provides goods or services* to a firm, these may be considered subsidies to the extent provision is made for less-than-adequate remuneration to the government or, if a purchase, the price provided more-than-adequate remuneration.

The second element is that a subsidy must be "specific" to a firm or industry if the government limits access to the assistance in law or in fact.[6] Assistance provided automatically, according to objective criteria or conditions established in law, is not specific. Setting or changing generally applicable tax rates will not constitute subsidies.

"Specificity" has two important dimensions: industrial and geographic. If a development program is restricted to certain industries or firms, it is specific, and potentially subject to discipline. Geographic specificity is more complex. With the exception of the "regional development" exemption described below, if a granting government directs subsidies to a certain region within its jurisdiction, those subsidies are specific. This can preclude state and local governments from offering subsidies for specific regions within their boundaries. To avoid the specificity condition, development programs must be "generally available" throughout the jurisdiction of a state or local government.

Trade Distortion

The GATT is concerned with only trade-distorting subsides. The new rules establish three categories of possible trade effects that define how a subsidy is to be treated: prohibited, actionable, and nonactionable. These categories are sometimes described as "red," "yellow," and "green," and the whole scheme is referred to as the "traffic light" approach.

Prohibited subsidies (red). Subsidies that are provided, explicitly or implicitly, to encourage exports or that are contingent on the use of domestic goods over imports are automatically trade-distorting and thus prohibited. These "export subsidies" were prohibited by the 1979 Subsidies Code, and these provisions remain substantially unchanged. If found as a result of a GATT subsidies investigation, prohibited subsidies must be removed.

Actionable subsidies (yellow). These are domestic subsidies against which remedies may be sought *if* they are shown to distort trade. This distinction is critical, because many economic development programs might be defined as subsidies under the rules but would only be of concern if they actually distort trade. However, as the rules have been written and are likely to be enforced, many programs that might not be considered trade distorting on a prima facie basis might still be caught in a GATT-authorized sanction.

. . . many economic development programs might be defined as subsidies under the rules but would only be of concern if they actually distort trade.

110

Trade distortion occurs if subsidized products

1. cause injury to a domestic industry when imported in competition with the products of that industry,
2. displace or impede the exports of another country from a third-country market, or
3. displace or impede imports from another country.

The latter two are new provisions. Subsidies have traditionally been a concern in the first case. For example, if subsidized lumber from Canada is imported into the United States, injuring U.S. lumber producers, the United States would have a complaint against the Canadian lumber imports. The second category means that complaints can now be raised if, for example, U.S. exports to Asia are displaced by subsidized exports from Europe. The third category means that complaints can now be raised if U.S. products receiving subsidies from U.S. governments (federal, state, or local) are able to underprice imports from other countries into the United States.

The Uruguay Round agreement adds a new concept of "presumed serious prejudice" as a special category of subsidies having a high likelihood of trade distortion. (This category has come to be called "deep yellow.") Presumed serious prejudice occurs when any of the following conditions are met: (1) subsidies total more than 5% of the value of a firm's or industry's output of the product subject to complaint, or (2) cover operating losses of a firm, or (3) forgive debts. In any of these cases, the burden of proof as to whether trade distortion has occurred moves from the complaining party to the defending (subsidizing) party. That is, if development programs result in presumed serious prejudice, the burden would fall on the state or local government to show that trade effects were *not* occurring, rather than the complaining country having to show that they were.

To be remediable, the value of all subsidies received by a firm or industry must generally be in excess of 1% of the value of the firm's or industry's sales: Subsidies received by different state and local governments may be cumulated to determine whether this 1% threshold is exceeded or whether the 5% threshold of presumed serious prejudice is exceeded. Thus it is not the total of subsidies provided by a single state or local government that is critical but the value of *all* subsidies received by an industry from governments anywhere in the United States.

. . . it is not the total of subsidies provided by a single state or local government that is critical but the value of *all* subsidies received by an industry from governments anywhere in the United States.

Nonactionable subsidies (green). This is an entirely new category in the GATT subsidy rules. The following subsidies will *generally* not be subject to discipline:

- Assistance with such research and development costs as personnel, equipment, buildings, and overhead, up to 75% of the costs of industrial research or 50% of the costs of "precompetitive development." Precompetitive research encompasses research up to the creation of the first product prototype. This covers assistance granted by public institutions of higher education, including activities typical of university research parks.
- Regional development programs "consistent with a general framework of assistance" that define eligible areas as "clearly contiguous with a definable economic and administrative identity" according to "neutral and objective criteria."[7] These criteria must include either (1) 85% or less of per capita income or household income or gross domestic product per capita compared to the granting jurisdiction as a whole or (2) an unemployment rate 10% higher than the unemployment rate in the granting region. This will be the critical element in permitting such programs as enterprise zones. Although most state or local programs stipulate either one or the other of income or unemployment in their eligibility criteria,[8] some enterprise-zone or similar programs may have to be modified to receive the "green light."
- Subsidies that pay up to 20% of the costs of meeting new environmental regulations through new equipment or production processes. These subsidies must be available only once and may not cover the cost of replacing or operating the assisted investment or any part of manufacturing costs. They must also be available to all firms adopting measures to comply with the new regulations.

111

These subsidies are nonactionable so long as they are notified to the World Trade Organization (the new name for the GATT institutions) before implementation (a 3-year period is provided to bring existing programs into compliance). Governments may choose not to notify the World Trade Organization on assistance for research and development, but if they do not, these subsidies will be actionable if they exceed the allowable percentages. Subsidies in the nonactionable category are generally not subject to remedies but may be subject to complaints if they cause "serious adverse effects" to an industry in an importing country.[9]

Provisions concerning the definition of nonactionable subsidies and presumed serious prejudice are to be in effect until 1999. At the end of that period, a GATT committee on subsidies will review operation of the provisions and consider whether they should be extended.

Remedies

If a trade-distorting subsidy is suspected, there are two options. If suspected injury from subsidized imports is the issue, the affected industry may request an investigation under national laws, and if injury to that industry is found, duties equal to the amount of the subsidy may be imposed on the imported goods to countervail (offset) the benefits of the subsidy. There is also the option in this case, or any other case in which subsidies are suspected of distorting trade, of initiating the dispute settlement process established by the new GATT rules.

As part of the new World Trade Organization, the GATT-1994 created a Dispute Settlement Body (DSB). The DSB can be used to settle disputes related to subsidies or any other trade-regulating agreements that are part of the new GATT. The general procedure in using the DSB is to first engage in consultations between the disputing parties; if this does not resolve the conflict, a panel of neutral trade law experts is established to hear the dispute. The panels' decisions regarding whether a subsidy exists and whether trade-distortion is occurring are generally binding on both parties. If subsidy-caused injury is found, the panel may direct removal of the subsidy or other steps to alleviate the harm done. If a party fails to comply with the panel recommendation, the panel may authorize trade retaliation, meaning the imposition of quotas, higher tariffs, or other restrictions on the products of the offending country that would not otherwise be permitted under GATT rules. In the past, these restrictions could only have been imposed on like goods, but the new rules permit "cross-retaliation"—that is, a reduction in trade rights on goods or services not directly involved in the dispute.

The agreement requires countries imposing countervailing duties to follow rules of procedure and evidence generally consistent with existing U.S. law. This should improve the accessibility and "transparency" (the trade policy term for a clearly defined process) of countervailing duty proceedings in countries that do not adhere to U.S. standards of administrative practice.

A critical element in determining whether subsidies will be subject to discipline is whether they cause injury.[10] The GATT-1994 adds language defining injury for the first time, drawn in large part from U.S. principles. Injury will generally be held to occur when, following receipt of the subsidy, a "significant increase" in the volume of subsidized imports occurs that results in or threatens to result in "significant price undercutting" or that prevents price increases. The rules call for a positive showing of the injury that occurs as a result of the subsidized imports.

An important exception to the GATT discipline on export subsidies, dating back to the origins of the GATT itself, applies to primary products, defined as "products of farm, forest, or fishery, or any mineral in its natural form or which has undergone only such processing as is customarily required to prepare it for marketing in substantial volume in international trade."[11] In the case of export subsidies, the GATT encourages members to "avoid" the use of subsidies but does not prohibit them. This permits the farm export subsidies extensively used by the United States, Europeans, and many other countries.

[GATT] permits the farm export subsidies extensively used by the United States, Europeans, and many other countries.

Farm subsidies were also a major issue in the GATT negotiations, since high levels of both domestic and export subsidies provided by the EC have created real problems for agricultural markets around the world. Limited reductions in both export and domestic subsidies on agriculture formed part of the final agreement, and there is some expectation that agricultural exports may

112

someday be brought within the general subsidies regime. However, for the foreseeable future, agricultural subsidies will not be subject to the same rules as are other subsidies.

2. THE U.S.-CANADA AND NORTH AMERICAN FREE TRADE AGREEMENTS

The increased specificity of the new provisions and their direct application to state and local development programs make the new GATT rules the most important element in the international order on subsidies. However, subsidies were also a central issue in the negotiations leading up to the U.S.-Canada Free Trade Agreement, and special provisions in that agreement, modified somewhat by NAFTA, create additional rules that may affect development programs.

The question of subsidized imports tended to dominate U.S.-Canadian trade relations during the 1980s. Disputes involving American industry claims of federal and provincial subsidies to the lumber and fish industries, two of Canada's major exporters, dominated Canadian headlines, and American unhappiness with the Canadians resulted in a tie vote in the Senate Finance Committee on the resolution authorizing the negotiations leading to the U.S.-Canada Free Trade Agreement, one vote shy of killing the entire deal before it ever got started. As a result, U.S. representatives in the U.S.-Canada Free Trade Agreement negotiations were instructed to persuade Canadians to reduce or eliminate their subsidies. Canadian negotiators, on the other hand, were instructed to protect the Canadian federal and provincial governments' rights to pursue regional development policies, while seeking limits on the American government's use of countervailing duties to, as Canadians saw it, "harass" Canadian imports.[12]

These positions were, to put it mildly, divergent. In large part, they reflected rather different views of the role of government in the economy,[13] and thus were not matters easily subject to negotiation. There was a decision early in the negotiations, as in most negotiations, to put the most contentious issues off to the latter stages and to deal with more tractable questions first. In the closing days a major effort to resolve the subsidies issue was made, but it proved too difficult. A resolution was necesary, and a "temporary" solution was found and incorporated as Chapter 19 of the agreement. This approach was also incorporated in Section 404 of NAFTA, which extended the approach to Mexico as well.

These chapters establish a new procedure for reviewing decisions on countervailing duty decisions in each country. In the United States, decisions to impose countervailing duties are normally appealed to a special Court of International Trade, but any such decisions involving Canada or Mexico are now referred to a special five-member panel. The panel reviews each decision to determine whether it was properly made in accordance with the law of the country applying the duty; if the panel finds the original decision was improperly made, its decision is binding.

Another special panel procedure was established to review changes in the laws of each country affecting countervailing duties. The review is to determine whether such changes contravene the GATT rules on countervailing measures or the general purpose of the free trade agreement, or whether the statutory changes served to overturn panel decisions regarding a specific countervailing duty case. If the special panel concludes that a statutory change violates any of these conditions, the other country can either make a similar change to its own laws or withdraw from the agreement. This provision reflected Canada's fear that even if a panel decided in its favor, Congress would simply change the law.

The U.S.-Canada agreement provided for these panel procedures to exist for a five-year period (which expired January 1, 1994), during which time the two countries were to negotiate another system of rules on subsidies and countervailing duties. An automatic two-year extension was included, after which either party could withdraw from the agreement. At the time this provision was agreed to, the Uruguay Round negotiations were under way; it was hoped that those negotiations might lead to more satisfactory arrangements for both countries. In fact, as discussed earlier, the Uruguay Round did result in a significant reform both in the definition of subsidies and in more rigorous rules on countervailing duties. At the same time, NAFTA made the procedural reforms respecting appeals permanent among the United States, Canada, and Mexico.

113

These two regional agreements do not raise the same substantive issues for state and local governments that the GATT agreement does, since they are procedural in nature. But the potential remains for effects on development programs depending on decisions by panels made under these agreements, which may differ from decisions of GATT panels. Although it seems unlikely, state and local development programs could find themselves subject to contradictory or at least inconsistent rulings from GATT panels and from NAFTA panels.

3. COMPLEXITIES

Substantially expanded as the GATT agreement on subsidies is, the true implications for state and local governments will become clear only after it begins to be interpreted and applied in specific cases. A brief review of key issues that have been raised about U.S. subsidies and countervailing duty practice may suggest a few areas in which controversy is likely to arise in implementing the new rules.

One key area will certainly be interpretation of the concept of specificity. While central to the new definition, it has been questioned by both scholars[14] and the U.S. Court of International Trade. In various cases, the court has issued such contradictory rulings on the subject that it has been called a "house divided."[15] The use of specificity as a key determinant of when a government expenditure is considered a subsidy came from American practice, in one of the defining cases of U.S. subsidy law. Significantly, the concept began with a case involving subsidies provided by a Canadian province, and the large financial incentives offered to the Michelin Tire Company of France to locate its first North American manufacturing facility in Nova Scotia.[16] In searching for a way to distinguish a domestic subsidy from an export subsidy, the Commerce Department zeroed in on the fact that the assistance in question was provided to only one firm and, thus, that trade distortion was likely. This was actually the first time the United States applied a countervailing duty to a domestic subsidy, and the case law was later enacted in U.S. law.[17]

The difficulties in this concept became clear in 1985 with the *Cabot* case, more commonly known as the "carbon black" case. American producers of carbon black, a petroleum-based raw material used in manufacturing, had challenged the Department of Commerce's decision that Petroleos Mexicanos' (PEMEX, the national oil company) provision of carbon black to the two Mexican producers at a price based on $2.00 per barrel oil when the going market price was $26.00 per barrel was not a subsidy. Commerce had reasoned that the lower price was not intended to benefit any particular firm, and so the lower price was generally available.[18] The International Trade Court held that although it might have been generally available in theory, the end result was still specific to the Mexican carbon black firms.[19]

The decision in the *Cabot* case became the basis for more confusion when lawyers for the American lumber industry spotted the decision and recognized that the International Trade Court had just reversed the Commerce Department on precisely the point on which the lumber industry had been rejected two years earlier in their first attempt to impose countervailing duties based on Canadian provinces' provision of low-cost timber to their own mills. In the 1983 case filed by the American lumber industry, alleging that low crown (public) land stumpage prices were a subsidy to the Canadian lumber industry (classified in the United States as Standard Industrial Classification [SIC] code 24), the Department of Commerce determined that low stumpage prices were not a subsidy because other industries, notably the pulp and paper industry (SIC 26), also benefited from the low prices.[20] Following the *Cabot* case in 1986, the Department of Commerce rejected its earlier logic and decided that the stumpage policies were targeted to the lumber industry.[21]

The issue raised by this case is this: What is an industry for purposes of determining general availability or specificity? Both the United States and Canada employ standard industrial classification systems as a taxonomy for describing and analyzing industries. The classification divides industries into four or more levels of detail, described by a code in which the number of digits signifies the level of detail. For example, in the American system, all manufacturing industries have codes beginning with "2" or "3"; all food-processing firms begin with "20," fish-processing firms with "203," and frozen-fish firms with "2031." Which of these is an "industry" within the

114

meaning of a specific subsidy? In practice, the U.S. Department of Commerce would define a subsidy available at the one-digit level (i.e., all manufacturing) as generally available; a subsidy received only by industries at the four-digit level would certainly not be considered generally available. The argument, therefore, will arise around the middle ground.

Another problem will likely arise in interpreting the language defining the value of loan subsidies as the difference between the government loan and "a comparable private loan which the firm can actually obtain in the market." In cases involving European steel producers, the U.S. position was that loans (or guarantees) made to creditworthy companies (those having a commercial credit rating) constitute a subsidy if they are below commercial rates and if the subsidy is equal to the difference between these and commercial rates. (Loans or guarantees made to an uncreditworthy company are considered to be the equivalent of an equity investment and treated as such.[22]) The view of the EC was that a loan or guarantee that has conditions attached to it is not a subsidy. The decreased cost of the loan is seen as offsetting the increased costs of the conditions. The EC's practice would largely offset the amount of the costs, greatly reducing the value of any subsidies. Although the U.S. Treasury Department (which had subsidy-determining responsibilities prior to 1979) took a similar view, Congress significantly narrowed the scope of offsets in the Trade Act of 1979[23]; only the costs of applying for assistance, losses in value resulting from delay in receiving assistance, or certain taxes levied on the export of the product now are permitted as offsets. These would usually be minor, so effectively, no offsets in the form of loan conditions are permitted under U.S. law.[24]

The new GATT rules may provide sufficient clarity to reduce the confusion of subsidies law, but odds are against it. The United States has had a countervailing duty statute since 1890, yet a hundred years later, a U.S.-Canada Free Trade Agreement panel reviewing a U.S. countervailing duty on pork imports from Canada could only complain that "[d]espite continuing efforts of [the Department of] Commerce and the courts to explain and clarify this issue . . . the standards for finding . . . specificity remain underdeveloped, opaque, contradictory, elusive, and unsatisfying."[25]

4. THE FUTURE FOR STATE AND LOCAL DEVELOPMENT PROGRAMS

The increasing support given for international trade development may well be considered "red-light" subsidies, whereas most other economic development programs that transfer resources directly to businesses are likely to fall within the yellow category.

Economic development programs are likely to come under scrutiny in all three categories under the new GATT code. The increasing support given for international trade development may well be considered "red-light" subsidies, whereas most other economic development programs that transfer resources directly to businesses are likely to fall within the yellow category. The green-light (permitted) category offers some leeway for development programs, but many existing programs may have to be changed to fall within the specific criteria spelled out in the GATT agreement. The extent of change necessary depends primarily on the extent to which products become subject to international trade disputes.

It may be argued that state and local government subsidies are small relative to those of other countries, and thus relatively unlikely to fall afoul of the thresholds defining subsidies or the requirement that subsidies actually cause injury to other countries' industries. There is some data to suggest this is the case, as Table 1 shows for national government subsidies (excluding tax concessions) in 1985. However, no systematic data on the size of expenditures and subsidies in state and local development programs in the United States exist, which makes the argument that these programs are too small to matter speculative at best, to say nothing of what this argument says about the effectiveness or value of the development programs.

Individually, state and local development programs are not likely to upset trade patterns. But if these programs become the subject of trade disputes, it almost certainly will not be on an individual basis. Since trade disputes involve industries, all programs affecting a given industry will be included in the dispute. This is specifically envisioned in the agreement provision that permits cumulation of all subsidies received by an industry in the subsidizing nation, and is entirely consistent with American past practice.

TABLE 1

Subsidies as a Percentage of Gross Domestic Product

Country	*Percentage*
Italy	3.43
France	3.01
Canada	2.48
United Kingdom	2.22
West Germany	2.01
Japan	1.15
United States	0.58

SOURCE: International Monetary Fund, *Issues and Developments in International Trade Policy*, International Monetary Fund Occasional Paper no. 66 (Washington, DC: International Monetary Fund, 1988).

In the dispute involving subsidies to the Canadian groundfish industry, a number of development programs offered by the provinces of Quebec, Nova Scotia, Newfoundland, New Brunswick, and Prince Edward Island were summed up and added to federal subsidies to reach the final countervailing duty figure.[26] Many of these programs were extremely small on an ad valorem basis, but when added together, they constituted nearly half of the total 4.7% value of subsidies found. Thus it is entirely possible for a very small state or local program to be mixed in with much larger similar programs in other jurisdictions when a dispute arises: A North Dakota program may be at risk because of another program in North Carolina.

Moreover, the cause of a dispute may not be another country's concern about state and local development programs per se but may arise out of American concerns about subsidies elsewhere. The international effort to impose disciplines on government subsidies has been driven largely by the United States, more specifically, by American industries that see themselves as competing, not with foreign countries, but with foreign governments. Congress has been receptive to these complaints, enacting laws to counter subsidized imports and insisting that reducing subsidies be a central element in major international trade policy negotiations. Americans lead the world, by far, in filing countervailing duty petitions; from 1981 to 1986, U.S. countervailing investigations accounted for between 60% and 100% of each year's investigations filed by the United States, Australia, Canada, Japan, and the EC.[27]

The American demand for ever-stricter discipline on subsidies has largely ignored the role that state and local governments have played in economic development efforts. The green-light categories in the Uruguay Round agreements were fiercely resisted by American industries, who considered these carve-outs as one more license for other countries to unfairly subsidize their products in competition with American goods. Indeed, state and local advisers to the negotiations had to insist on such exceptions being kept in order for the U.S. government to agree to them as a part of the price of the deep-yellow category.[28]

It may be reasonably expected, given past U.S. interest in attacking foreign subsidies and the increased standard of review in the Uruguay Round agreement, that American industries will continue to aggressively pursue subsidies in other countries. The result may be the ensnaring of state and local programs. Precisely this scenario occurred in 1991-1993 in the beer dispute with Canada, which did not involve subsidies but demonstrates what can happen when such programs are made the subject of trade disputes.

Canadian provinces have had laws since the 1930s requiring that beer sold in the province must be brewed there, the purpose being to ensure that jobs in the brewing industry remain in each province. Two small Wisconsin-based breweries, Heileman's and Stroh's, complained about these restrictions in a case brought before a GATT panel. Recognizing that they were on shaky ground, the Canadians brought a counterclaim against a variety of state and local liquor laws, including requirements that beer be imported in common carrier trucks (to provide audit trails for tax enforcement), as well as tax exemptions for microbreweries and brew-pubs. Their hope was to have a bargaining chip to use if the Canadians lost the case. In fact, GATT panels ruled against both the Canadian provincial restrictions and the American state laws, despite the trade effects of

The American demand for ever-stricter discipline on subsidies has largely ignored the role that state and local governments have played in economic development efforts. . . . It may be reasonably expected, given past U.S. interest in attacking foreign subsidies and the increased standard of review in the Uruguay Round agreement, that American industries will continue to aggressively pursue subsidies in other countries. The result may be the ensnaring of state and local programs.

the Canadian restrictions being more severe and obvious. States are now in the process of changing laws and procedures to comply with the panel ruling.

These cases, known as Beer I (the U.S. complaint) and Beer II (the Canadian complaint), were the first GATT cases involving the activities of state and local governments, and they set a troubling precedent. They show how vulnerable state and local activities may be to retaliatory actions initiated by other countries as well as how GATT panels can equate relatively innocuous state and local programs with more egregious violations of GATT principles, requiring the same response in terms of changes to laws in both cases.

On balance, therefore, it seems likely that state and local development programs will be exposed not only in theory but also in practice to challenge under the newly expanded GATT rules on subsidies. One may contemplate three possible futures for state and local development programs:

1. *Voluntary conformity.* State and local development programs are significantly modified over the next few years to bring them into strict conformity with the new rules.
2. *Involuntary conformity.* Aggressive efforts to remove subsidies initiated by major U.S. trading partners or in retaliation for U.S. efforts result in a widespread attack on state and local subsidies and widespread changes in development programs, partly as a result of GATT or national countervailing duty decisions, partly as other governments change their programs as a defense against future attacks.
3. *Sporadic conformity.* Development programs are targeted in occasional complaints, and elements of development programs are found in violation of GATT rules and subjected to countervailing duties or GATT-panel requirements for modification or removal.

Of these three scenarios, the third seems most likely. The first seems the least likely given the political and organizational investments in existing programs and the speculative nature of sanctions under the GATT. Governments are unlikely to undertake wholesale changes in their programs on a prospective basis. The second scenario is certainly possible, although its probability is unknown. The third scenario must thus be seen as most likely at this point, inasmuch as the principal effect of the new international rules on subsidies is to greatly increase the uncertainties faced by those creating and managing economic development programs.

There are other important issues raised by these expanded disciplines that may affect development programs. One is whether the new international discipline on subsidies will have any effect on the often cutthroat competition between state and local governments. Competition, in the form of tax reductions and other incentives, has been a constant element in the drive toward ever-expanding state and local development programs, and many commentators have questioned both the effectiveness and the efficiency of these competitive pressures.[29] It is clear that the increasing pressure on domestic subsidies is meant "as a check against internal political process" that tends to result in increasing subsidies to business."[30] If governments choose to conform to the new rules, it may reduce the competition, but the pressures for development will likely remain. Development programs may thus continue to be an arena of competition, if within new boundaries.

Beyond traditional economic development programs, other state and local policies with a differential impact on the location or performance of businesses may be considered as potentially suspect subsidies. The most likely may be those involving environmental protection. This can be seen in the claims that free trade with Mexico would result in unfair competitive advantages for Mexican firms because of allegedly lower environmental standards. Regulation as a subsidy has also been identified by several writers as a potential issue in conflicts over the appropriateness of government intervention.[31] Both NAFTA and the GATT agreements contain sections on "technical barriers to trade" that seek to limit the ability of governments to use environmental or other standards as disguised barriers to trade. NAFTA goes considerably beyond this in the environmental side agreement, which essentially gives each party the right to challenge other parties for failing to enforce environmental laws. The connection between trade and the environment, which was tentatively formed in NAFTA (and, to a lesser extent, in the Uruguay Round agreements), is widely considered to be the next major frontier in international trade policy. State and local governments will barely get used to the new rules on subsidies before being faced with a new set of international rules defining, or seeking to define, the proper relationship between regulation and development.

State and local development policies have recognized the realities of the global economy for more than a decade. But among those realities is the necessity of rules to govern trade, rules that will now reach state and local governments in unprecedented ways and whose ultimate effects are unforeseeable. What is likely is that these rules will become a major influence on the evolution of economic development policies in the next decade.

NOTES

1. Agreement of Subsidies and Countervailing Measures in U.S. Congress House of Representatives *Uruguay Round Agreements Act*. House Document 103-826. 103d Cong., 2d sess., 1994. Washington, DC: U.S. Government Printing Office; U.S. Congress House of Representatives *North American Free Trade Agreement: Texts of Agreement, Implementing Bill, Statement of Administrative Action, and Required Supporting Statements*. House Document 103-159. 103d Cong., 1st sess., 1993. Washington, DC: U.S. Government Printing Office; U.S. Congress House of Representatives *The United States-Canada Free Trade Agreement*. House Document 100-216. 100th Cong., 2d sess., 1988. Washington, DC: U.S. Government Printing Office.

2. National Governors' Association, *Going Global: A Governor's Guide to International Trade* (Washington, DC: National Governors' Association, 1992).

3. J. Barceló, "A History of GATT Unfair Trade Remedy Law—A Confusion of Purposes," *World Economy* 14 (September 1991): 311-34.

4. G. C. Hufbauer and J. S. Erb, *Subsidies in International Trade* (Washington, DC: Institute for International Economics, 1984); J. Barceló, "Subsidies, Countervailing Duties, and Antidumping after the Tokyo Round," *Cornell International Law Journal* 13 (1980): 257-88.

5. Agreement on Subsidies and Countervailing Measures, Article 14, paragraph (c).

6. The concept of specificity was drawn from U.S. Commerce Department procedures in deciding countervailing duty cases, procedures upheld in U.S. law in *Carlisle Tire and Rubber Company v United States*, 564 F. Supp. 834 (Court of International Trade 1983).

7. Agreement on Subsidies and Countervailing Measures, Article 8, paragraph 8.2 (b) (i-ii).

8. U.S. Department of Housing and Urban Development, *State Designated Enterprise Zones: 10 Case Studies* (Washington, DC: U.S. Department of Housing and Urban Development, 1986).

9. Firms may receive more than one of these green-light subsidies. A provision restricting firms to only one of the classes of subsidies was deleted from an earlier draft. See A. Anderson, "An Analysis of the Proposed Subsidies Code Procedures in the Dunkel Text of the GATT Uruguay Round: The Canadian Exporter's Case," *Journal of World Trade Law-Economics-Public Policy* 27 (June 1993): 70-100.

10. The agreement also permits complaints under the general category of "nullification and impairment" of benefits otherwise due under the GATT agreements.

11. General Agreements on Tariffs and Trade, note to Article XVI, section B.

12. A. Rugman and A. Anderson, "A Fishy Business: The Abuse of American Trade Law in the Atlantic Groundfish Case of 1985-86," *Canadian Public Policy* 12 (1987): 152-64; A. Rugman and A. Anderson, *Administered Protection in America* (New York: Methuen, 1987); M. Lefebvre, *U.S. Protectionism: The Lumber Issue* (Montreal: National Bank of Canada Department of Economic Analysis, 1986).

13. S. M. Lipset, *Continental Divide: The Values and Institutions of the United States and Canada* (London: Routledge, 1990).

14. P. Jameson, "Recent International Trade Commission Practice regarding the Material Injury Standard: A Critique," *Law and Policy in International Business* 18 (1986): 517-77; J. Panzarella, "Is the Specificity Test Generally Applicable?" *Law and Policy in International Business* 18 (1986): 417-20; C. Lehmann, "The Definition of Domestic Subsidy under U.S. Countervailing Duty Law," *Texas International Law Journal* 22 (1987): 50-95.

15. M. Sussman, "Countervailing Duties and the Specificity Test: An Alternative Approach to the Definition of Bounty or Grant," *Law and Policy in International Business* 18 (1986): 475-515.

16. R. V. Guido and M. Morrone, "The Michelin Decision: A Possible New Direction for U.S. Countervailing Duty Laws," *Law and Policy in International Business* 6 (1974): 237-66.

17. Tariff Act of 1930, Section 775, 19 U.S.C. Section 1677[5].

18. United States International Trade Administration, "Preliminary CVD Results: Carbon Black from Mexico," *Federal Register* 51, no. 75 (1986): 13269ff, microfiche.

19. *Cabot Corp v. United States* is at 620 F. Supp. 722 (Court of International Trade 1985). The decision was later codified in U.S. law in the 1988 Omnibus Trade and Competitiveness Act, 19 U.S.C. Section 1677(5)(B).

20. United States International Trade Administration, "Final Negative Countervailing Duty Determination: Softwood Lumber Products from Canada," *Federal Register* 51, no. 105 (1983), microfiche.

21. United States International Trade Administration, "Preliminary Affirmative CVD Determination: Softwood Lumber Products from Canada," *Federal Register* (1986), microfiche.

22. C. Barshefsky, A. Mattice, and W. Martin, "Government Equity Participation in State-Owned Enterprises: An Analysis of the Carbon Steel Countervailing Duty Cases," *Law and Policy in International Business* 14 (1983): 1101-58; J. Davidow, "Hegemony with a Vengeance: U.S. Trade Law Attacks on Foreign Government Equity Participation in Exporting Enterprises," *Columbia Journal of Transnational Law* 24 (1986): 279-99.

23. Trade Agreements Act of 1979, Public Law 96-39. Washington, DC: U.S. Government Printing Office.

24. J. H. Bello and J. J. Talbot, "Foreign Government Subsidy Reductions: When Should the Commerce Department Take Them Into Account?" *George Washington Journal of International Law and Economics* 19 (1985): 109-21.

25. U.S.-Canada Free Trade Agreement Binational Review Panel In Re Certain Fresh Pork Products from Canada, September 28, 1990 (Washington, DC: Office of the United States Trade Representative).

26. United States International Trade Administration, "Preliminary Affirmative CVD Determination: Softwood Lumber," 24568.

27. G. Hufbauer, "A View from the Forest," in *Subsidies and Countervailing Measures: Critical Issues for the Uruguay Round*, World Bank Discussion Papers, no. 55, ed. B. Balassa (Washington, DC: World Bank, 1989), pp. 11-26.

28. Intergovernmental Policy Advisory Committee, *Report of the United States Trade Representative's Intergovernmental Policy Advisory Committee on the Agreements Reached in the Uruguay Round of Multilateral Trade Negotiations* (Washington, DC: Office of the United States Trade Representative, 1994).

29. D. Netzer, "An Evaluation of Interjurisdictional Competition through Economic Development Incentives," in *Competition among State and Local Governments*, ed. O. Kenyon and J. Kincaid (Washington, DC: Urban Institute, 1991).

30. Barcel6, "A History of GATT."

31. E. J. Rowbotham, "Dumping and Subsidies: Their Potential Effectiveness for Achieving Sustainable Development in North America," *Journal of World Trade* 27 (1993): 145-73.

28

KYKLOS, Vol. 49 – 1996 – Fasc. 1, 47 - 68

LIPPINCO
DOCUMEN

WHARTON REPROGRAPHICS

Are Discriminatory Procurement Policies Motivated by Protectionism?

Albert Breton and Pierre Salmon*

I. PUZZLES CREATED BY THE DOMINANT INTERPRETATION OF PREFERENTIAL PURCHASING

When purchasing goods and services, governments often discriminate against foreign suppliers. The available estimates of the size of the bias are precarious, but the existence of a bias is well-documented for both central and sub-central (i.e., regional and local) governments[1]. In this paper, we take the discrimination as fact. There is, in addition, a general agreement that discriminatory practices should be eliminated. That prescription has been largely endorsed, as a matter of principle, in the agreements that concluded GATT's Tokyo and Uruguay Rounds, in a series of Directives from the European Union (formerly European Community), in the decisions of the European Court of Justice, and in the comments of almost all economics and law scholarship concerned with the issue. We also note that governments, even though collectively the source of the prescription, have been severally reluctant to implement it.

---

* University of Toronto, Department of Economics, 150 St. George Street, Toronto, Ontario M5S 1A1, Canada; Université de Bourgogne, Faculté de Science économique, 4 boulevard Gabriel, 21000 Dijon, France. For helpful comments and suggestions, we are grateful to Jean Magnan de Bornier, Michel Mougeot, Martin Paldam, participants at the European Public Choice Society Meetings, European Economic Association Congress and Séminaire Jean Bénard (Université de Paris I Panthéon-Sorbonne), as well as to an anonymous referee. We are of course solely responsible for the shortcomings. We are also grateful to the Lynde and Harry Bradley Foundation for its support.

[1] For some estimates at the level of central governments see, e.g., Baldwin (1970), Lowinger (1976), Atkins (1988) and the references therein. At the level of sub-central governments see, e.g., Breton (1967), Finsinger (1988), Dalpé (1992) and, again, the references therein. The bias is typically measured by contrasting public sector with overall purchases of foreign (or out-of-jurisdiction) goods. In European countries, for instance, according to the Atkins Report, imports represented in 1986-88, on average, 22 percent of GNP but only 2 percent of public purchasing. For some evidence that figures such as these underestimate public purchasing of foreign goods and services, see Hartley and Uttley (1994).

Our main concern is not normative. The prescription is interesting to us because it reflects the quasi-unanimous belief that preferential procurement policies are exclusively motivated by protectionism. We therefore ask: Are discriminatory procurement policies always motivated by protectionism? Although it seems to be perfectly obvious, a positive answer to this question is puzzling in several ways. The main puzzle stems from a point made several years ago by Baldwin and Richardson (1972) which has not, however, received sufficient attention. Baldwin and Richardson showed that when imported goods and domestically produced goods are perfect substitutes, and when government purchases of these goods are only a fraction of domestic supply, discriminatory government procurement is completely ineffective as a protectionist device. Because it plays an important role in the development of our own argument, we reproduce the essential of their basic proposition.

Figure 1

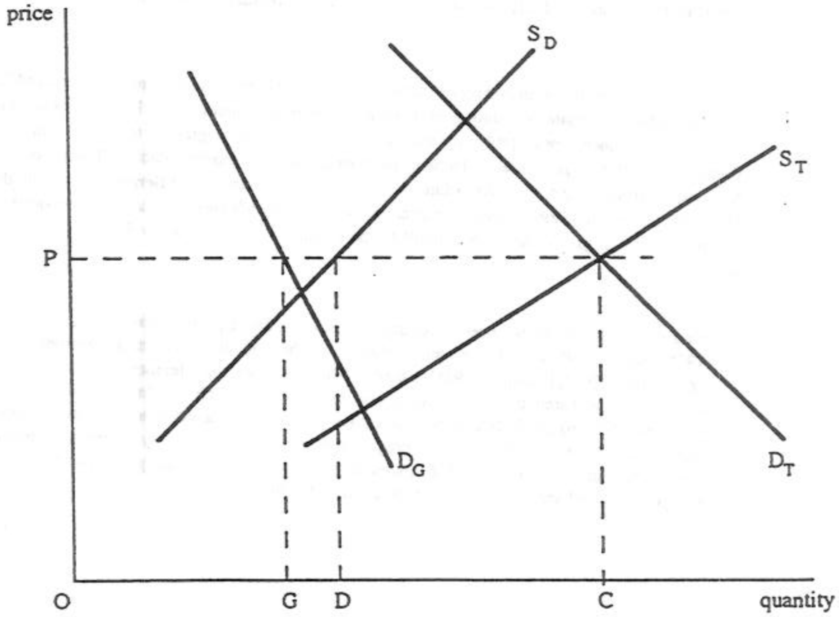


Figure 1 is a simplified version of Baldwin's Figure 6.1 (1984, p. 602) and is similar to the figure used by Baldwin and Richardson (1972). For a good whose characteristics are as stated above, let D_G be the demand curve of the public sector, D_T the total demand curve in the country (public plus private sector demand), S_D the supply curve of domestic producers, and S_T the total supply curve in the country (domestic supply plus imports). In the absence of discrimination, the price of the good is equal to OP, total consumption equal to OC, domestic supply to OD, government purchases to OG and imports to DC. Even if the government decides to buy exclusively from domestic suppliers, the values taken by these variables do not change. The upward pressure on price resulting from the government's exclusive patronage of domestic suppliers is offset by an increase in the purchase of imports by the private sector. In Baldwin's opinion, this result is not an artefact of modelling but is highly relevant in the real world. He writes:

> 'Since government purchases of most non-military items represent only a fraction of the domestic supply of these items, the preceding analysis suggests that government preferential purchasing policies are generally ineffective in raising the price received by domestic suppliers. Nor will they affect domestic employment, production, profits, or growth' (p. 603).

Baldwin and Richardson's analysis was limited to competitive industries but it has been extended to oligopolistic settings by Miyagiwa (1991). Although the results are ambiguous in some cases, the proposition that discriminatory procurement policies are ineffective is generally confirmed – and indeed reinforced (imports increase) – under various assumptions about substitutability, the precise form given to discrimination, and returns to scale. When the conditions for Baldwin and Richardson's and Miyagiwa's results hold, even exogenous tariff or pressure groups explanations cannot be used to support the interpretation of discrimination as a manifestation of protectionism. The only possibility that remains is that governments and/or voters are ignorant – an assumption that we should use sparingly at least for persistent and widespread behaviours.

When government demand is not a fraction of domestic supply, but instead exceeds that supply, discrimination raises domestic prices and production and reduces imports. But the price of imports to private consumers falls and the private sector imports more[2]. As a consequence, the aggregate effect of discrimination on imports is likely to be small (see especially Richardson 1972). This means that if discrimination is to be used as an instrument of protectionism, it is likely to be expensive in terms of procurement costs and yield little in terms of protection – it is likely to be cost-ineffective. If we now assume that central and sub-central governments, not unlike business enterprise, are competitive, we must then suppose that these governments supply goods and services to their citizens at the lowest possible unit cost. A well-known mechanism that drives governments to minimise costs is based on the mobility of people and capital (Tiebout 1956, Stigler 1972). Another less well-known mechanism is based on the assumption that voters assess the performance of their government – in respect of the volume and quality of the goods and services which it makes available to them at specified prices – by comparing that performance with the performance of governments in corresponding jurisdictions elsewhere, in respect of the same variables (see Salmon 1987, Breton 1996). With electoral competition in each jurisdiction and uncertainty on the part of office-holders about the assessments of voters, the comparison of performances provides these office-holders with an incentive to perform as best as they can on all dimensions and, therefore, to minimise costs. Recent empirical work (Besley and Case 1995) as well as casual observation of budgetary discussion and of budget-making at all levels of government suggest that governments are indeed constrained both in the raising of funds and in the allocation of these funds between competing requirements. On the basis of this analysis, we consider a policy likely to be cost-ineffective as also puzzling[3].

---

[2] This formulation, borrowed from Baldwin (1984), implies that the government will pay for a homogeneous good a price which is higher than the price paid by private buyers, albeit, contrary to Herander (1982), not from the same sellers.

A third puzzle is the following. In every country there are many public sector buyers rather than a single unitary actor called 'the government'. Each one of them is submitted to budgetary constraints and to competition. Why, insofar as they remain free – or even exhorted – not to discriminate against foreign suppliers, would these public sector buyers discriminate more against foreign suppliers than against out-of-jurisdiction suppliers in general? In the case of a state or local government, preferential treatment of own-jurisdiction suppliers can be understood within the dominant paradigm, but favourable treatment of domestic albeit out-of-jurisdiction suppliers is more difficult to explain. Still, that is what we observe in a number of countries. For example, a number of states in the United States impose upon themselves 'buy-American' constraints, in addition to the constraints that are imposed on them by the federal government (see, e.g., Peterson 1980, p. 336).

---

[3] One could argue that the same kind of puzzle would be created by the assumption that office-holders maximise their rents and thus have no reason to buy above minimum cost even when not constrained by competition. But office-holders may have other objectives than maximising their (pecuniary) rents. They may, for instance, seek 'a quiet life'. The information provided to voters by competition of the relative performance variety impedes this kind of behaviour. For a similar argument in the context of agency costs within firms, see Holmstrom and Tirole (1989, pp. 95-96).

The view that discrimination is obviously motivated by protectionism rests on an implicit model of what government procurement would be in the absence of protectionist motivations. This model is itself based on a conception of exchange in which transactions between buyers and sellers are embodied in complete contracts that are legally enforceable at zero cost. Because of this assumption, any difference between the long-run market-clearing price and marginal cost is *prima facie* evidence of imperfections in competition of which discrimination and protectionism are examples. Under the influence of Armen Alchian, Benjamin Klein, Keith Leffler and others, this conception has been modified to permit a reinterpretation of practices such as price discrimination, resale price maintenance, exclusive territories, franchising and other manifestations of vertical restraint which is not only more realistic but which is consistent with resource-economising behaviour on the part of economic agents.

Incomplete contracts are an essential ingredient in this conceptual transformation. They embody transactions in which the parties can be symmetrically informed, can monitor each other (at the limit at zero cost), but which, because of the high cost of ascertaining all relevant circumstances, cannot be monitored by a third party[4]. Performance is then said to be non-verifiable, which implies that incomplete contracts cannot be legally enforced. Still, as long as net gains to trade are assumed to exist, it must be presumed that rational actors will want to exploit these gains and that they will, as a consequence, strive to invent an effective enforcement mechanism. On the assumption of repeat business, such a mechanism will exist if a buyer (say) is capable of creating, for the benefit of the seller, a stream of quasi-rents whose present worth is at least as large as the value to the seller of defaulting on the contract. In the theory of efficiency wages for example, employees are offered the prospect of a flow of quasi-rents – a premium over the market clearing wage – if they behave according to the wishes of the employer and are dismissed if they behave otherwise, which implies losing the quasi-rents. This logic is basic to our analysis of governmental procurements. Indeed, we attempt to show that preferential purchasing, by making the creation of a flow of quasi-rents feasible, can be an effective way of securing seller performance[5].

---

[4] As a consequence, the problem of incomplete contracting can arise not only in the context of long-run agreements but also when the relationship is modelled as a series of short-run contracts. This point is stressed by Williamson (1991) in his critical comments on the assumptions made in Fudenberg, Holmstrom and Milgrom (1990).

The objective of the paper is a relatively simple one. It is to construct a model capable of explaining why a practice such as preferential purchasing by governments continues to be observed in the face of a nearly universal consensus, endorsed by these governments themselves, that these preferences are to be condemned. In other words, our purpose is to provide a rationale derived from, and consistent with, recent developments in the theory of contracts to explain the long-term persistence of preferential procurement policies by governments.

II. BUYING AS A QUASI-HIERARCHICAL RELATIONSHIP

In their well-known paper on the organisational forms of production structures, Alchian and Demsetz (1972) argued that there is no more authority in firms than there is in markets, in which, they claimed also, there is none. This view that authority is not an essential characteristic of firms has been widely criticised. One objection is that most contracts within firms are incomplete. But Alchian and Demsetz's proposition provides us with an important insight, namely that relations in markets are not in all dimensions different from those in enterprises. In markets as in firms, quasi-rents are created by one party to the benefit of others as inducements to deliver what has been promised. One application of this enforcement mechanism is to vertical relationships between firms, namely to the relation of industrial buyers with their suppliers and to the relation of firms with their franchisees or retailers (Klein and Murphy 1988). In this context, the enforcement mechanism which is often the most powerful and the most robust to changing circumstances is also the simplest: it relies exclusively on the unilateral creation of a prospective stream of quasi-rents in exchange for performance. To borrow a phrase coined by Arthur Stinchcombe (1985), the buyer-seller relationship, then, 'simulates the operations of hierarchies'.

Notwithstanding national accounting conventions, governments are seldom (final) consumers of the goods and services they buy. For the great bulk of their activities they are, indeed, better conceived as producers and suppliers of goods and services and, therefore, as buyers of inputs. Furthermore, they are not single monolithic providers, but generally operate in competitive environments. As suggested earlier, there is competition, based on relative performance, between the multiplicity of governments that make up governmental systems. In addition, within each government, there is competition between different branches, departments and agencies. It is, therefore, reasonable to assume that governmental agencies will require and expect from their own suppliers, at least in part, the same kind of services that business firms require from their suppliers. It is also reasonable to suppose, consequently, that the analyses which have been developed to understand buyer-seller relationships in regard to business enterprises have some relevance for the analysis of similar relationships in the context of governmental agencies.

---

[5] In the volume edited by Leitzel and Tirole (1993), several contributions analyse government procurement under the assumption that contracts are incomplete, and the importance of quasi-rents and repeat business as a way to ensure supplier performance is stressed in Julio Rothenberg's brief comment (pp. 72-73); discrimination against out-of-jurisdiction suppliers is not envisaged, though.

When we turn our attention to the literature that addresses these concerns – especially to the work on marketing – we note that it begins by stressing the importance of the services which accompany the delivery of goods to industrial firms. The following list of the more salient of these services is adapted from that literature (see for example Noordewier, John and Nevin 1990, and the references therein). It includes:

1. Information
The provision, prior to signing the contract, of adequate information about the goods and services to be supplied;

2. Client's needs
A recognition of the particular needs of the buyer and a willingness to adjust to these needs;

3. Maintenance
The capacity and willingness to provide after-sale maintenance services, to deal with follow-up problems, to deliver replacement parts, etc.;

4. Purchasing
The ease with which the buyer can deal with invoicing, accounting, stock-keeping and other problems related to the acquisition or purchase itself;

5. Assistance
The capacity and willingness to install the purchased input, to train personnel in using it, to smooth out the psychological costs of using the input, etc.;

6. Exclusivity and secrecy
Buyers may want that one or more inputs be of exclusive design and not made available to other suppliers, at least for a period of time, and they may also require that the exact nature of the inputs be kept secret. The willingness and

ability of the seller to oblige is, then, a valuable complementary service to buyers;

7. Monitoring
The ease with which the behaviour of sellers can be monitored not only in respect of the inputs supplied, but also in respect of other actions which could be a source of public embarrassment to buyers;

8. Flexibility
The willingness and capacity of sellers to forgive payment delays, alterations in the conditions of supply while delivery is in progress, trespass on agreed rules or conditions, cancellations, short notice acceleration of delivery, etc.[6]

Most purchases will not require all of these services. For some inputs, it is easy to verify quantity, quality and other characteristics of services; for others, buyers can rely on the reputation of sellers. But, for some of the services just listed – and, no doubt, for others – neither third-party enforcement nor the reputation of sellers will do. For these, the prospect of repeat business and of quasi-rents is essential to insure performance.

To proceed, we assume without loss of generality that a particular government agency, G_x, buys the same bundle of goods every period, say every year. For that bundle of goods, G_x is one among M buyers. Over a single period, together with the goods themselves, G_x buys from supplier S_i (one among N potential suppliers) the provision of a number of complementary services. We make no assumption about the acquisition procedure used, which can range from public bidding to free negotiation, except that we suppose that G_x selects the proposal that it finds the most advantageous. The delivery of the goods and their quality, as well as the provision of some associated services, are verifiable; the other services are not. To simplify, we further assume that all transactions take place at the beginning of the period (thus eliminating discounting problems) and according to the following timing:

- G_x selects S_i and signs a contract with that supplier;
- S_i delivers the bundle of goods and the associated verifiable services;
- G_x controls their quality;
- G_x pays the agreed price (for the whole package);
- S_i provides the unverifiable services;
- G_x observes the unverifiable services.

---

[6] We note that in the case of 'flexibility', third-party enforcement is ruled out almost by definition: as it is conceived in the text, flexibility typically consists in giving-up third-party enforcement.

The first thing that G_x does at the beginning of the period therefore is to select S_i. That supplier is the successful bidder because (if 'most advantageous' is equated with 'least costly') the total cost of the bundle of goods plus the verifiable and unverifiable services it offers is lower than that of the other firms in the pool of potential suppliers. If we let g be the cost of the bundle of goods plus that of the verifiable services, u the cost of the unverifiable services, q the quasi-rent and m the monopoly-oligopoly rent (see below), the total cost of procurement is $g + u + c$, where $c = q + m$. To simplify, and again without loss of generality, we assume throughout that u is the same for all suppliers; we also assume that once the size of the pool of potential suppliers (N) has been determined (Sections III and IV), the excess costs c are also the same for all firms. The selection of the successful bidder is, therefore, made on the basis of total cost or of g alone because, once N is known, both lead to the same selection.

In the remainder of this section and in the two that follow, we propose a model of the determination of the pool of successful suppliers and of the level of c (= q + m) and, therefore, a model of total cost that G_x must pay to purchase a bundle of goods (plus accompanying services). We begin by recognising that S_i may not provide the unverifiable services expected by G_x. To avoid this outcome, G_x includes in the price a quasi-rent (a premium) in addition to the total cost of the bundle of goods plus the verifiable and unverifiable services. Associated with the prospect of repeat business, this may prove effective in inducing S_i to perform. S_i will compare V_p, the remuneration expected in the case of performance and V_{np}, the remuneration expected in the case of non-performance. It will decide to perform if $V_p \geq V_{np}$.

Let us consider V_p first. It is assumed that in future periods G_x will continue to buy the same unverifiable services (along with the same goods and verifiable services) from some supplier. But even if satisfied with the unverifiable services provided by S_i, G_x by no means commits itself to again select S_i in these future periods. Since costs vary from firm to firm – and, for each firm, from period to period – in any future period another supplier may make a more advantageous offer to G_x than S_i could profitably make. G_x would then turn to this new supplier (this does not imply that G_x would not, at a still later date, contract anew with S_i)[7].

What then is the probability that S_i will be selected by G_x in each future period when S_i provides the unverifiable services it has promised to deliver? That question, it should be stressed, will be central to S_i's decision as to whether it should perform or shirk. In what follows, we assume that this probability is related to the size of the pool of potential suppliers of the goods and services under consideration. To be more specific, assume that S_i is risk-neutral and that, in period 1, it forms the following expectations: first, that the number of firms (N) able to supply the relevant inputs is constant over time; second, that the probability for any of these firms (including itself) to make the most favourable offer in each future period is equal to 1/N; and, third, that both the cost of the unverifiable services and the size of the quasi-rent remain constant over time. If r is the rate of interest and q the quasi-rent received in each period, then the present (expected) value of the future stream of quasi-rents (i.e., V_p), assuming an infinite time horizon, is equal to q/rN.

---

[7] To simplify, we assume that only one supplier is awarded a contract in each period.

Suppose that S_i does not provide the unverifiable services. The simplest and, at the same time, the limiting case is when S_i assumes that this will lead to a permanent discontinuance of every kind of relation with G_x. Then, future quasi-rents are completely lost[8], and the gain from reneging on the promise to deliver unverifiable services is simply the cost of these services, that is, u. In that simple case, $V_{np} = u$ and the condition that $V_p \geq V_{np}$ reduces to:

$$q/rN \geq u$$

or, assuming strict equality:

$$q = urN \tag{1}$$

This expression tells us that as N increases continuously, the size of the quasi-rent that the buyer must create to prevent shirking by the seller also rises.

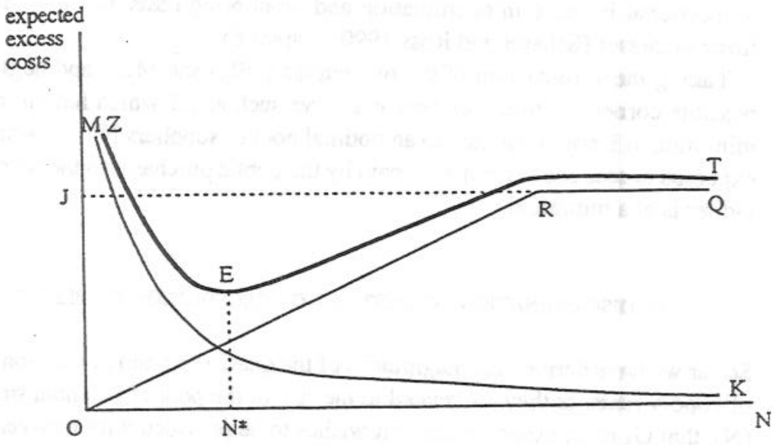
---

[8] Because S_i cannot really form the expectation that G_x will refuse to do business with it (S_i) under all circumstances, we should merely assume that, when G_x is dissatisfied with S_i's performance, it places a lower value on S_i's promise to perform in the future (see Laffont and Tirole 1993, p. 227, for a related discussion). Suppliers will remain uncertain about the exact effects of their actions but, with expected quasi-rents that are sufficiently attractive, they will nonetheless be induced to please G_x as much as they can. (This is exactly what happens in a hierarchy, as already argued by Albert Breton and Ronald Wintrobe (1982), when subordinates provide informal services to their superiors in the expectation of an undisclosed future reward). Under these circumstances, the precise calculation of the quasi-rent provided in the text is no longer valid. Indeed, because S_i is now excluded from repeat business over a finite number of periods rather than forever and because it remains necessary that the expected stream of quasi-rents be superior to the given gain from shirking (u), the value of the quasi-rent per period (q) must be higher. The main qualitative result of the analysis holds however: reducing the number of potential suppliers reduces the quasi-rent that must be paid by G_x and thus makes possible the obtaining, on 'reasonable terms', of the desired unverifiable services.

For example, with 3 potential suppliers and an interest rate equal to 5 percent, the quasi-rent must be equal to 15 percent of the cost of the unverifiable services to discourage shirking; with 100 suppliers and the same rate of interest, the quasi-rent must be 5 times the cost of the unverifiable services to achieve the same end.

It should be clear however that the buyer will not pay any sum whatsoever for the unverifiable services and, consequently, that the linear relationship between N and q (equation (1)) is upwardly bounded by a constraint defined either by the monetary equivalent of the utility derived from the services or by the money cost of the best alternative available, the smaller of the two being shown as OJ in *Figure 2*. In other words, that *Figure* 2's ORQ curve, passing through the origin and having slope ur (equation (1)) has a discontinuity at R (height equal to OJ), the point at which the value of the unverifiable services to the buyer, or the cost of an alternative arrangement, is equal to the cost of the quasi-rent.

Figure 2



Following a long tradition which to our knowledge begins with Edgeworth, it is reasonable to assume that, as the number of enterprises in an industry increases from one to a very large number – at the theoretical limit, infinity – the monopoly-oligopoly rents accruing to the firms and, therefore, the price in excess of the competitive price which the buyer must pay, decrease giving rise to a curve such as MK in *Figure 2*. Two features of that curve should be noted. One, the height of point M is determined by the equilibrium monopoly price – the monopoly rent when there is only one firm in the industry. Two, the reduction in the monopoly-oligopoly rent brought about by the addition of one firm will be smaller when there are many firms than when there are a few. This is depicted in *Figure 2* by the shape of the MK curve: falling rapidly at first and more slowly afterwards. Two rationales can be given for that shape of MK. First, under, certain types of interactions – for example, under Bertrand competition – the competitive result is obtained after the entry of only one additional firm. Under other types of interaction, the competitive outcome is obtained or approached less rapidly. Being uncertain about the exact weight that should be given to the various types of interaction, it must be assumed that on the average MK will fall more rapidly when there are only a few firms than when there are many. Second, collusion and the probability of its incidence will be larger the smaller the number of firms – a result that can be ascribed to the more than proportional increase in coordination and monitoring costs as the number of firms increases (Scherer and Ross 1990, chapter 8).

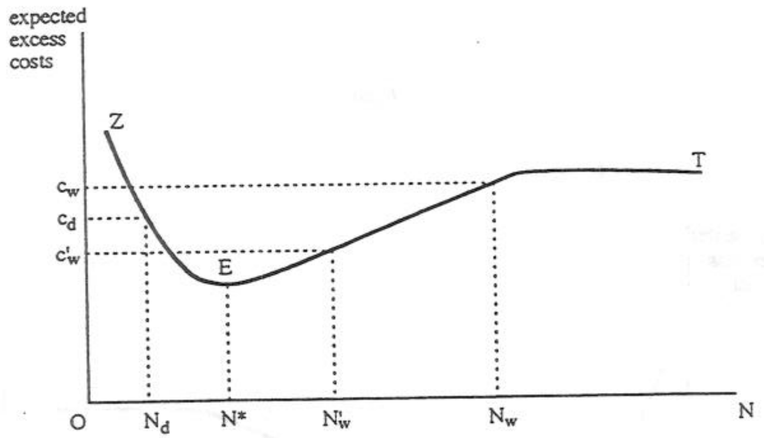
Taking the vertical sum of the two curves (ORQ and MK), and neglecting possible corner solutions, we obtain a curve such as ZT which has an interior minimum at E corresponding to an optimal pool of suppliers N* – at which the expected excess costs ($c = q + m$) paid by the public purchaser to the successful bidder is at a minimum.

III. DISCRIMINATION AGAINST OUT-OF-JURISDICTION SUPPLIERS

So far we have derived the magnitudes of the quasi-rents and of the monopoly-oligopoly rents, as they are related to the size of the pool of potential suppliers (N), that G_x must expect to pay if it wishes to be provided with non-verifiable services by S_i. Assuming that G_x's decisions are made on the basis of ZT, G_x's only decision, if N is given, is whether to purchase or not to purchase the unverifiable services, that is, for any N, whether the associated expected cost of the quasi-rent is lower or higher than OJ.

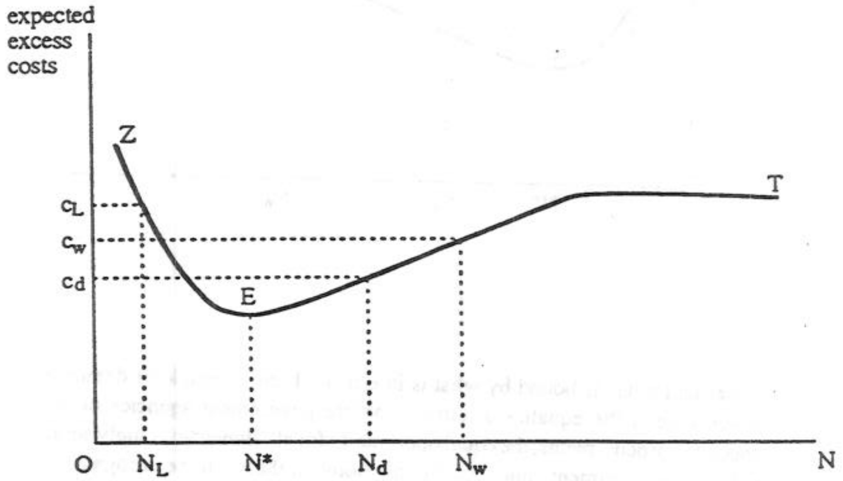
In the remainder of the paper, we examine alternative ways that G_x may use to alter the size of the pool of potential suppliers so as to minimise the expected excess costs of non-verifiable services. A basic consideration that we will have to keep in mind is that G_x is a governmental or public decision-maker and subjected to constraints that are not binding for private purchasers that are in other respect similarly positioned.

Figure 3



In particular, G_x is bound by what is in effect a basic principle of democracy which asserts the equality of citizens before government agencies or, as it applies to procurements, the equal opportunity for all 'domestic' supply sources to bid for government contracts. At any moment, there will be derogations of various sorts from that principle but we neglect these in what immediately follows. It is this principle which will lead G_x, if it wants to reduce the pool of potential suppliers, to discriminate against 'foreign' or, more precisely, out-of-jurisdiction suppliers. To understand how discrimination under the principle of democracy operates consider *Figure 3*. Assume first that the total (world) pool of suppliers is equal to N'_w of which N_d are 'domestic' enterprises. Under these circumstances, the expected costs of unverifiable services to G_x are less when no restrictions are placed on the size of N; that is, because $c'_w < c_d$ – the expected excess costs associated with N'_w and N_d respectively – there is no need to discriminate. N'_w is the size of the pool of suppliers that make expected excess costs as small as possible. If, however, the world pool of suppliers is N_w of which the domestic component is still N_d, the expected excess costs are c_w and c_d, and $c_d < c_w$. Under these circumstances, G_x will want to reduce N and, because of the principle of democracy, will do so by discriminating against foreign or out-of-jurisdiction enterprises[9].

Figure 4



---

[9] Within the theory of auctions, some researchers have formally demonstrated that, in a context of competitive bidding, discrimination against foreign suppliers may significantly lower procurement costs. For two different analyses along that line, see McAfee and McMillan (1989) and Mougeot and Naegelen (1992). Both assume complete contracting.

The reader will recall our discussion, identified above as the 'third puzzle', of the situation which is sometimes encountered of local jurisdictions which discriminate more against foreign suppliers than against other out-of-jurisdiction enterprises, as happens when American state, Canadian provincial, or French regional authorities decide to abide by a 'buy American', 'buy Canadian' or a 'buy French' policy. That such a policy may be efficient can be seen by reference to *Figure 4* in which N_w is again the worldwide pool of suppliers, while N_d and N_L are respectively its 'domestic' (American, Canadian, or French) and 'local' (state, provincial or regional) components. As drawn, *Figure 4* makes clear that 'local' jurisdictions can reduce the expected excess costs of non-verifiable services by adopting 'buy-domestic' policies. We note that the discrimination enacted by the 'local' authorities when they adopt such policies is based on a differentiation between 'foreign' (genuinely foreign plus domestic out-of jurisdiction) sources of supply but not between their 'own' suppliers. Thus, it does not violate the democratic principle discussed earlier.

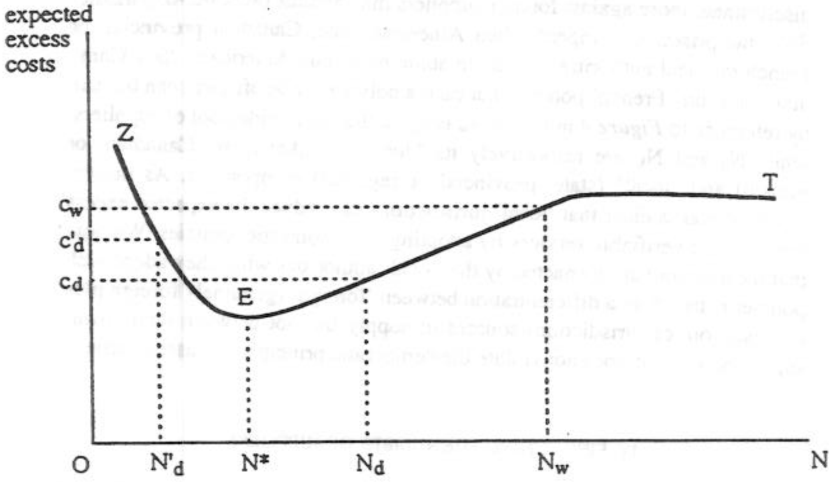
IV. FINE-TUNING THE NUMBER OF SUPPLIERS

In the discussion of the last section, we were concerned with the analysis of discrimination under the equalitarian principle of democracy. G_x, the public decision-maker, was therefore limited to actions that would restrict or increase the pool of potential suppliers through discrimination that was respectful of that principle. It is time to recognise that, except by accident, the pool of suppliers obtained through that instrumentality will not be N^*, the size of the pool that would minimise the expected excess costs of the unverifiable services. We should therefore expect that government purchasers will seek to reduce – or, possibly, increase – the pool of suppliers by making use of different instruments. This section is concerned with the analysis of the operation of a subset of these instruments and begins by focusing on instruments used to reduce N.

Consider the situation depicted in *Figure 5* when the world pool of suppliers is N_w of which N_d are domestic firms. G_x will first discriminate against foreign suppliers, but will in addition wish to reduce the pool of 'domestic' firms to attain N^*. How can this objective be achieved? We suggest that this is done by implementing rules of transparency (such as open bidding and public disclosure of sealed bids) and regulatory controls (like bookkeeping requirements, proof of internal monitoring, and certification of compliance) which in the words of Kovacik (1992, p. 34)

> 'discourage otherwise eligible firms from taking risks and bearing costs uniquely associated with regulatory compliance in public contracting.'

Figure 5



Kovacik, though he decries the regulatory measures as inefficient, argues that they act as 'barriers to entry in government procurement'. We agree, but contend that, as long as $N_d > N^*$ in *Figure 5*, the barriers promote efficiency by allowing government purchasers to reduce the excess costs of buying non-verifiable services. Kovacik's general line of reasoning raises an interesting problem. He points out that the Competition in Contracting Act of 1984

> 'required each executive agency to establish a 'competition advocate' who would promote ...'full and open competition' and 'challenge barriers to such competition' '(p. 39).

The pursuit of these objectives, Kovacik argues, was frustrated by the expansion of regulatory controls which, in effect, led to the opposite result (p. 30).

On the general precept, formulated earlier, that it is inadvisable to presume persistent and widespread ignorance and irrationality, we would argue, on the contrary, that the Act of 1984 was always intended by Congress to generate barriers that would induce purchasers to economise on scarce resources. In addition, we insist that the Act had to be formulated in a language consistent with the equalitarian principle of democracy as well as acceptable to the era's *Zeitgeist*.

We must now consider briefly the situation depicted in *Figure 5* when the world pool of suppliers is N_w, but the domestic pool is $N'_d < N^*$. To reduce the excess cost of buying unverifiable services, the public purchaser will want to increase the number of domestic suppliers. This situation is not likely to be encountered for the bulk of the goods purchased by governments, but can still be important, measured by dollars of expenditures, for certain goods such as military hardware, nuclear installations and transportation equipment such as trains, aeroplanes, buses and subway cars when public purchasers prefer to have three or, at least, two potential suppliers instead of one. Under these circumstances, we expect governments to subsidise either directly or indirectly the development of one or two domestic enterprises.

Figure 5 is not very useful to analyse this case because the public subsidies will displace the ZT curve upward and to the right – the subsidy will increase the unit cost of non-verifiable services and the increase in N will move N_w to the right, because N'_d plus the additional firms are part of the world total – but the general idea that the excess cost of non-verifiable services can in some cases be lowered by the public subsidisation of new ventures is, however, easily intuitioned with the help of the figure which makes clear that an increase in the number of domestic suppliers will lead to a downward movement on the new curve, signalling a reduction in the excess costs of unverifiable services which can, in some cases, exceed the cost of the subsidy.

V. IMPLICATIONS

Some of the implications of the foregoing analysis have already been noted. In this section, after a brief discussion of how the approach we are proposing can in principle be tested, we examine two additional implications of the model, both of which pertain to the relationship between the size of jurisdictions and the incidence of non-protectionist procurement policies.

It would be a simple matter to test the hypothesis if information on the following variables (measured across categories or classes of goods and services) was readily available:

- the level of discrimination and/or the speed (or lack thereof) at which discrimination is reduced;
- the degree to which the conditions for the Baldwin-Richardson-Miyagiwa ineffectiveness proposition are satisfied;
- the relative importance of unverifiable services;
- the effect of discrimination on the excess cost of procurement (as measured vertically in *Figures 3*, *4* and *5*) or, more precisely, on the two components of that cost, namely the monopoly-oligopoly rent and the quasi-rent.

Then, for any class of goods, if discrimination in government procurement is significant and has resisted liberalisation efforts, if the Baldwin-Richardson-Miyagiwa conditions are satisfied, if unverifiable services are important, and if the reduction in the pool of suppliers caused by the exclusion of out-of-jurisdiction supply has a negative effect on the quasi-rent which is larger than the reduction's positive effect on the monopoly-oligopoly rent, we would have to conclude that, with respect to these goods, our analysis is capable of accounting for the facts and that, as a consequence public procurement is not motivated by protectionism.

Turning our attention to the first of the two implications we wish to discuss, we note that, if for some bundle of goods, $c_d < c_w$, but the number of potential suppliers within the jurisdiction is large such that N_d lies to the right of N^* (as in *Figures 4* or *5*), we expect to observe, as noted earlier, not only discrimination against out-of-jurisdiction sources of supply but also the implementation of policies aimed at reducing N_d. In addition to what has been said about this in Section IV, the model also suggests that, *ceteris paribus*, the flow of regulatory measures whose indirect effect is to reduce N_d would increase as the size of markets increases and with it the number of potential suppliers. A *ceteris paribus* comparison between large and small jurisdictions for example, should reveal that the conditions for the Baldwin-Richardson-Miyagiwa ineffectiveness proposition will be satisfied more often and that N_d will be large more frequently in larger jurisdictions than in the smaller ones. As a consequence, we should expect both the incidence of non-protectionist public procurement and the incidence of measures designed to reduce N_d to be greater in larger than in smaller jurisdictions.

The second implication adds another dimension – the presence of economies of scale in production – to the size of the jurisdiction. The liberalisation of procurement policies – the elimination or reduction of discrimination and preferential treatment – currently implemented in Europe (see, e.g., Hartley and Uttley 1994) is mainly motivated by the gains expected from the exploitation of scale economies. The quantitative estimates presented in the Atkins Report (1988) indicate that these economies constitute the main part of the benefit expected from liberalisation. As a result of discrimination, and thus of fragmentation of markets, the size of many European firms in sectors in which government procurement is important or even dominant (e.g., defence or telecommunications) – or the scale of production for particular goods (e.g., a particular fighter aircraft) – is small relative to the optimum. Liberalisation of procurement *within* Europe is expected to generate a reorganisation of European industries that will entail the survival, in each of these sectors, of only a few large firms. According to the Atkins Report, this will considerably reduce the cost of procurement for all European governments (for defence procurement, see some impressive estimates in Fontanel and Smith 1991).

Our hypothesis can account for the willingness of member states to accept a liberalisation of government procurements which leads to diminished procurement costs for goods and verifiable services and which, at the same time, maintains a small number of suppliers and therefore a lower cost for unverifiable services. It can also account for the reluctance of the same member states to liberalise on a wider scale – that is, on one in which the number of potential suppliers would be much larger. Hence, their attachment to *préférence communautaire* and their timidity in implementing liberalisation in a GATT setting – a phenomenon which international relations specialists find puzzling (see Grieco 1990).

V. CONCLUSION

Discriminatory governmental procurement policies are habitually taken to be *prima facie* evidence of protectionism. Without denying that discrimination may sometimes be motivated by protectionism, we have suggested a hypothesis based on the idea that discrimination may be an efficient method of dealing with the problems posed by the existence of incomplete contracts, namely contracts whose terms, while observable by the signing parties, cannot be verified by third parties and are, therefore, legally unenforceable. Discrimination is, in that framework, a way of assisting in the creation of a stream of quasi-rents to the benefit of suppliers, which, if they decide to shirk on their contractual commitments, they will lose. Discrimination is then an instrument that serves to make incomplete contracts enforceable. Granted that our hypothesis has some validity, it becomes important to design empirical work in such a way as to permit identification of those instances when procurement policies are protectionist and those when they are used to enhance overall efficiency. Though we have not considered normative issues, it is also clear that the hypothesis has far-reaching policy implications which future work will have to draw out.

REFERENCES

Alchian, Armen A. and Harold Demsetz (1972). Production, Information Costs, and Economic Organization, *American Economic Review*. 62: 777-795.

Atkins Report (1988). *Research on the 'Cost of Non-Europe', Basic Findings, Volume 5, Parts A and B: The 'Cost of Non-Europe' in Public-Sector Procurement*. Luxembourg: Commission of the European Communities.

Baldwin, Robert E. (1970). *Nontariff Distorsions of International Trade*. Washington, D.C.: Brookings.

Baldwin, Robert E. (1984). Trade Policies in Developed Countries, in: Ronald W. Jones and Peter B. Kenen (eds.), *Handbook of International Economics*. Amsterdam: North-Holland: 571-619.

Baldwin, Robert E. and J. David Richardson (1972). Government Purchasing Policies, Other NTB's, and the International Monetary Crisis, in: H.E. English and Keith A.J. Hay (eds.), *Obstacles to Trade in the Pacific Area: Proceedings of the Fourth Pacific Trade Development Conference*. Ottawa: Carleton School of International Affairs.

Besley, Timothy and Anne Case (1995). Incumbent Behavior: Vote-Seeking, Tax-Setting, and Yardstick Competition, *American Economic Review*. 85: 25-45.

Breton, Albert (1967). *Discriminatory Government Policies in Federal Countries*. Montreal: Private Planning Association of Canada.

Breton, Albert (1996). *Competitive Governments: An Economic Theory of Politics and Public Finance*. New York: Cambridge University Press.

Breton, Albert and Ronald Wintrobe (1982). *The Logic of Bureaucratic Conduct*. Cambridge: Cambridge University Press.

Dalpé, Robert (1992). Les marchés publics québecois et canadiens: la définition des politiques d'achat dans le contexte d'une économie ouverte, *Politiques et Management Public*. 10: 21-36.

Finsinger, Jörg (1988). Non-Competitive and Protectionist Government Purchasing Behavior, *European Economic Review*. 32: 69-80.

Fontanel, Jacques, and Ron Smith (1991). A European Defence Union? *Economic Policy*. 6: 393-424.

Fudenberg, Drew, Bengt Holmstrom and Paul Milgrom (1990). Short-Term Contracts and Long-Term Agency Relationship, *Journal of Economic Theory*. 51: 1-31.

Grieco, Joseph M. (1990). *Cooperation Among Nations: Europe, America, and Non-Tariff Barriers to Trade*. Ithaca and London: Cornell University Press.

Hartley, K. and M.R.H. Uttley (1994). The Single European Market and Public Procurement Policy: The Case of the United Kingdom, *Public Procurement Law Review*. 3: 114-125.

Herander, Mark (1982). The Impact of Government Price Discrimination and its Equivalence with the Tariff, *Weltwirtschaftliches Archiv*. 118: 525-545.

Holmström, Bengt R. and Jean Tirole (1989). The Theory of the Firm, in: R. Schmalensee and R.D. Willig (eds.), *Handbook of Industrial Organization*. Amsterdam: Elsevier Science Publishers. 1: 61-133.

Klein, Benjamin and Kevin M. Murphy (1988). Vertical Restraints as Contract Enforcement Mechanisms, *Journal of Law and Economics*. 31: 265-297.

Kovacik, William E. (1992). Regulatory Controls as Barriers to Entry in Government Procurement, *Policy Sciences*. 25: 29-42.

Laffont, Jean-Jacques, and Jean Tirole (1993). *A Theory of Incentives in Procurement and Regulation*. Cambridge, Mass.: The MIT Press.

Leitzel, Jim, and Jean Tirole (eds.) (1993). *Incentives in Procurement Contracting*. Boulder: Westview Press.

Lowinger, Thomas C. (1976). Discrimination in Government Procurement of Foreign Goods in the U.S. and Western Europe, *Southern Economic Journal*. 42: 451-460.

McAfee, R. Preston and John McMillan (1989). Government Procurement and International Trade, *Journal of International Economics*. 26: 291-308.

Miyagiwa, Kaz (1991). Oligopoly and Discriminatory Government Procurement Policy, *American Economic Review*. 81: 1320-1328.

Mougeot, Michel and Florence Naegelen (1992). Marchés publics et intégration européenne: le rôle des coûts d'entrée, in: Gilles Bertin and André Raynauld (eds.), *L'intégration économique en Europe et en Amérique du Nord / Economic integration in Europe and North America*. Paris: Editions Clément Juglar: 221-243.

Noordewier, Thomas G., George John and John R. Nevin (1990). Performance Outcomes of Purchasing Arrangements in Industrial Buyer-Vendor Relationships, *Journal of Marketing*. 54: 80-93.

Peterson, Deborah J. (1980). The Trade Agreement Act of 1979: The Agreement on Government Procurement, *Journal of International Law and Economics*. 14: 321-348.

Richardson, J. David (1972). The Subsidy Aspects of a 'Buy American' Policy in Government Purchasing, in: *The Economics of Federal Subsidy Programs: A Compendium of Papers Submitted to the Joint Economic Committee. Congress of the United States. Part 2 – International Subsidies*. Washington D.C.: U.S. Government Printing Office: 220-243.

Salmon, Pierre (1987). Decentralisation as an Incentive Scheme, *Oxford Review of Economic Policy*. 3: 24-43.

Scherer, F.M., and David Ross (1990). *Industrial Market Structure and Economic Performance*, 3rd. edition. Boston: Houghton Mifflin.

Stigler, George J. (1972). Economic Competition and Political Competition, *Public Choice*. 13: 91-106.

Stinchcombe, Arthur L. (1985). Contracts as Hierarchical Documents, in: A.L. Stinchcombe and C. Heimer (eds.), *Organization Theory and Project Management*. Oslo: Norwegian University Press.

Tiebout, Charles M. (1956). A Pure Theory of Local Expenditures, *Journal of Political Economy*. 64: 416-424.

Williamson, Oliver E. (1991). Economic Institutions: Spontaneous and Intentional Governance, *Journal of Law, Economics, & Organization*. 7, Special Issue: 159-187.

SUMMARY

When purchasing goods and services, governments often discriminate in favour of domestic suppliers. It is widely assumed that such behaviour is motivated by protectionism. Although this interpretation is sometimes valid, it is also puzzling. After reviewing some of the puzzles, the paper proposes an alternative explanation of preferential procurement based on the assumption that governmental buyers want to purchase goods and services at minimum cost, but must do this in a context in which, because of the presence of unverifiable services, contracts are necessarily incomplete. The paper argues that preferential purchasing can guarantee the efficient delivery of these unverifiable services.

ZUSAMMENFASSUNG

Ausgehend von einer einfachen Prämisse schlägt dieser Aufsatz einen Rahmen vor, mit Hilfe dessen unterschiedliche Korruptionsniveaus unter Regierungen, ihren Unterabteilungen als auch über die Zeit erklärt werden können. Die einfache Prämisse ist: Bei Korruption gibt es fast immer auch Verlierer, die je nachdem Gegenmassnahmen ergreifen, um ihre Verluste zu reduzieren. Die

Anreize, Gegenmassnahmen zu ergreifen, hängen von zwei Gruppen von Faktoren ab: globalen und spezifischen. Globale Faktoren, wie zum Beispiel Menschen-, politische und Eigentumsrechte, Ausbildung, Einkommensniveaus und Einkommensverteilung, haben allgemeine Auswirkungen auf die Fähigkeit, Gegenmassnahmen zu ergreifen. Unterschiedliche Korruptionsniveaus unter Ländern und über die Zeit können unter anderem mit diesen Faktoren erklärt werden. Spezifische Faktoren, wie beispielsweise die Korruptionsart, Anzahl Verlierer und Höhe ihrer Verluste, können zwischen den Regierungsstellen innerhalb des gleichen Landes variieren und deshalb deren verschiedene Korruptionshöhen erklären.

RÉSUMÉ

Dans leurs achats de biens et services, les Etats discriminent souvent en faveur des producteurs nationaux. On attribue généralement ce comportement à une motivation protectionniste. Cette interprétation est parfois valide, mais elle est aussi problématique. Après avoir exposé certaines des interrogations qu'elle suscite, le texte propose une autre explication de la discrimination dans les achats publics, fondée sur l'hypothèse que les acheteurs publics veulent se procurer des biens et services au coût minimum mais, en raison de la présence de services non-vérifiables, doivent le faire dans un contexte de contrats incomplets. L'article essaie de montrer que la discrimination peut garantir une fourniture efficace de ces services non-vérifiables.

Global Trade Policy

LIPPINCOTT
DOCUMENT DELIV

The World Trade Organisation: Watershed Innovation or Cautious Small Step Forward?

John H. Jackson

1. INTRODUCTION

ON 1 January 1995, a new international organisation came into being. Depending on one's perspective, the World Trade Organisation (WTO) is either a modest enhancement of the General Agreement on Tariffs and Trade (GATT) which preceded it, or a watershed moment for the institutions of world economic relations embodied in the Bretton Woods system.

For twelve years, over 100 nations of the world have been participating in the largest and most complex negotiation concerning international economics in history (some would say the largest and most complex negotiation ever). Launched formally under GATT auspices in Punta del Este, Uruguay in September 1986 (after some years of preparation), the Uruguay Round negotiating results were formally signed at Marrakesh, Morocco on 15 April 1994, and ratified by a sufficient number of nations to bring those results into force on 1 January 1995. The results are embodied in a 'document' of some 26,000 pages, most of which are detailed schedules of tariff, services trade, and other concessions. For the first time the GATT system includes major agreements on trade in services, and on trade related intellectual property questions. Some 50 other portions of the agreement address subjects as diverse as anti-dumping, agriculture trade, subsidies, technical standards, textiles, and customs valuation. Included in this agreement are two important institutional measures: the Charter for a WTO, and a new set of dispute settlement procedures, both designed to assist in the effective implementation of the substantive rules established in the agreements.

2. THE GATT AND ITS BIRTH DEFECTS

The GATT was never intended to be an 'organisation'. It was negotiated in 1947

JOHN H. JACKSON is Hessel E. Yntema Professor of Law at the University of Michigan.

© Blackwell Publishers Ltd. 1995, 108 Cowley Road, Oxford OX4 1JF, UK

WHARTON REPROGRAPHICS

30

and 1948, at the same time as negotiators prepared a Charter for an ITO (International Trade Organisation). The GATT was to be a multilateral trade and tariff agreement, which would depend for its organisational context and secretariat services, on the ITO, but the ITO never came into being because the United States Congress would not approve it. The GATT did not need US Congressional approval because it was negotiated under advance authority granted to the President in the 1945 extension of the Reciprocal Trade Agreements Act (the first such Act was 1934). Because the ITO failed to materialise, the GATT gradually became the institution for international government cooperation on trade matters. In fact, however, the GATT never definitively came into force; instead it was legally applied by a Protocol of Provisional Application originally designed to last only until the ITO came into force.

Despite this inauspicious beginning, the GATT has been remarkably successful during its nearly five decades of history. Partly this is because of ingenious and pragmatic leadership in the GATT, particularly in its early years, as the GATT struggled to fill the gap left by the ITO failure. As decades passed, however, there was recognition that the GATT system was being increasingly challenged by the changing conditions of international economic activity, including the greater 'interdependence' of national economies, and the growth in trade of services. Concern developed that the GATT was too handicapped to play the needed role of complementing the Bretton Woods system as the 'third leg', alongside the IMF and World Bank. Problems and 'birth defects' included:

- Provisional application and Grandfather rights exceptions embraced by the Protocol of Provisional Application.
- Ambiguity about the powers of the Contracting Parties to make certain decisions.
- Ambiguity regarding the waiver authority and risks of misuse.
- Murky legal status leading to misunderstanding by the public, media, and even government officials.
- Certain defects in the dispute settlement procedures.
- Lack of institutional provisions generally, so constant improvisation was necessary.

3. THE WTO

a. Genesis of the WTO

The 1986 Ministerial declaration of Punta del Este, containing the agenda and objectives for the Uruguay Round (UR) negotiation, did not include any explicit call for a new charter or organisation, although it did establish negotiating groups including one for dispute settlement, and another for 'Functioning of the GATT System', (FOGS). Despite this hesitancy, however, by 1990 there was considerable discussion of the need for an improved organisational structure for effective implementation of the Uruguay Round results. In December 1991, the Uruguay Round Negotiators led by the GATT Director General, Arthur Dunkel, prepared and released a draft text of treaty clauses which covered the entire UR negotiation results up to that point, with indications of work yet to do. This was an important project with many implications. Included in this draft (for the first time), was a tentative draft of a new Charter for an organisation — an MTO or Multilateral Trade Organisation. This draft had a number of flaws, recognised by the US government and others, but through hard work the negotiators were able to revise the draft and iron out most of the flaws. In the December 1993 near final draft, the new organisation was retitled the WTO — World Trade Organisation. This Charter is included in the treaty embodying the results of the Uruguay Round, along with a major text of new procedures and rules for dispute settlement.

b. What is the WTO?

With the new WTO Charter, the Uruguay Round results should provide a better institutional structure to fill the gap left in the Bretton Woods structure. Several general characteristics are noteworthy.

First, the charter for the WTO can be described as a 'mini-Charter'. It is devoted to the institutional and procedural structure that will facilitate and in some cases be necessary for effective implementation of the substantive rules that have been negotiated in the Uruguay Round. The WTO is not an ITO (the 1948 ITO draft Charter which never came into force). The WTO Charter text itself does not include substantive rules, but it incorporates the substantive agreements resulting from the Uruguay Round into 'annexes'. In many cases the criticism aimed at the WTO during the implementation debates was really criticism aimed at some of the substantive provisions of the Uruguay Round results, and should not be considered a criticism of the WTO institutional Charter.

Second, the WTO essentially will continue the GATT institutional ideas and many of its practices, in a form better understood by the public, media, government officials and lawyers. To some extent, a number of the GATT 'birth defects' are overcome in the WTO. The WTO Charter (XVI:I) expressly states the intention to be guided by GATT 'decisions, procedures and customary practices' to the extent feasible. The practice of consensus is better defined and for the first time becomes a legal procedure in some important decisions, rather than just a practice.

Third, the WTO structure offers some important advantages for assisting the effective implementation of the Uruguay Round. For example, a 'new GATT 1994' is created to supercede the 'old GATT'. This procedure avoids the constraints of the amending clause of the old GATT which might make it quite difficult to bring the Uruguay Round into legal force. At the same time, the WTO ties together the various texts developed in the Uruguay Round and reinforces the 'single package' idea of the negotiators, namely, that countries accepting the Uruguay Round must accept the entire package (with a few exceptions). No longer will the Tokyo Round approach of side codes, resulting in 'GATT a la Carte' be the norm.

The WTO Charter establishes (for the first time) the basic explicit legal authority for a Secretariat, a Director-General and staff. It does this in a way similar to many other international organisations, and it also adds the obligation for nations to avoid interfering with the officials of the organisation.

Another important aspect of the WTO structure is that it facilitates the extension of the institutional structure to the new subjects negotiated in the Uruguay Round, particularly services and intellectual property. Without some kind of legal mechanism such as the WTO, this would have been quite difficult to do since the GATT itself only applies to goods. The new GATT structure separates the institutional concepts from the substantive rules. The GATT 1994 will remain a substantive agreement (with many of the amendments and improvements developed throughout its history, including the Uruguay Round.) The WTO has a broader context. Similarly, the WTO will be able to apply a unified dispute settlement mechanism, and the Trade Policy Review Mechanism to all of the subjects of the Uruguay Round, for all nations who become members.

Fourth, the WTO Charter offers considerably better opportunities for the future evolution and development of the institutional structure for international trade cooperation. Even though the WTO Charter is minimalist, the fact that there is provision for explicit legal status and the traditional organisational structure, helps in this regard. With the WTO focusing on the institutional side, it also offers more flexibility for future inclusion of new negotiated rules or measures which can assist nations to face the constantly emerging problems of world economics.

c. *The Legal Structure of the WTO*

The WTO charter is confined to institutional measures, but the charter explicitly outlines four important annexes which technically contain hundreds of pages of substantive rules. The annex structure is important and the different annexes have different purposes and different legal impacts (see Appendix A).

Annex 1: This contains the large texts, termed 'Multilateral Trade Agreements,' which comprise the bulk of the UR results. All these are 'mandatory,' in the sense that these texts impose binding obligations on all members of the WTO. This reinforces the single package idea of the negotiators, departing from the Tokyo Round approach of 'pick and choose' side texts, or 'GATT a la carte.' The Annex 1 texts include:

Annex 1A: GATT 1994, the revised and all-inclusive GATT agreement with related agreements or 'Codes', and the vast 'schedules of concessions' that make up the large bulk of pages in the official treaty text. The schedules for each of the major trading countries, the US, Japan, and the European Union, each constitute a volume of printed tariff listings.

Annex 1B: GATS, is the General Agreement on Trade in Services, with its schedules of specific commitments and Annexes.

Annex 1C: TRIPS, is the Agreement on Trade Related Intellectual Property measures.

Annex 2: This contains the dispute settlement rules, which are obligatory for all members, and which form (for the first time) an integrated and unified dispute settlement mechanism covering the WTO Charter, the agreements listed in Annex 1, Annex 2, and made available for agreements in Annex 4.

Annex 3: This contains the TPRM — Trade Policy Review Mechanism, (in existence since 1988) by which the WTO will review the overall trade policies of each member on a periodic and regular basis, and report on those policies. The approach is not supposed to be 'legalistic,' and questions of consistency with WTO and Annex obligations are not the focus; rather the focus is on transparency and the general impact of the trade policies, both on the country being examined and on its trading partners.

Annex 4: This contains four agreements which are 'optional,' and termed 'Plurilateral Agreements.' This is a slight departure from the single package ideal, but the agreements included tend to deal with subject matter which concerns a small number of countries, or is more 'hortatory' in nature. Clearly this annex, to which addition may be made, provides some important flexibility for the Organisation to evolve and redirect its attention and institutional support for new subjects that may emerge as important during the next few decades. The agreements currently included in Annex 4 cover trade in civil aircraft, government procurement, dairy products, and bovine meat.

Annex 1A is by far the largest and contains the GATT 1994, which is essentially the old GATT as modified by amendments, the Tokyo Round 'codes' as renegotiated in the Uruguay Round, and some new Uruguay Round agreements. In addition to the GATT 1994, Annex 1A includes

agriculture,
sanitary and phytosanitary measures,
textiles and clothing,
technical barriers to trade,
trade-related investment measures,

Article VI (anti-dumping),
customs valuation,
preshipment inspection,
rules of origin,
import licensing,
subsidies and countervailing measures, and
safeguards,

plus a series of 'understandings' which further modify the GATT, and some Ministerial 'decisions and declarations.' The tariff schedules (most of the 26,000 pages), are technically part of this Annex. The relationship of many of these agreements listed above to GATT 1994 is somewhat murky.

Two of these agreements concern what are probably the most contentious of the 'rules of conduct' clauses of the GATT, namely Anti-dumping and Subsidies/Countervailing. One other concerns product standards (technical barriers) also addressed in the Sanitary and Phytosanitary agreement. As noted before, the legal relationship of these various GATT additions to the core GATT agreement itself is not always clear.

As impressive as the Uruguay Round results are, there clearly are a number of 'leftover' issues which the WTO system will need to address during forthcoming years, in addition to overseeing a satisfactory implementation of the UR results. The descriptions above hint at some of these, and other issues can be named. Together these include (1) enhancing and extending liberalisation of trade in agricultural products; (2) future extensive negotiations on services; (3) further elaboration of the rules on subsidies; (4) further market access efforts; (5) further negotiations in the context of trade related investment measures; (6) further negotiations concerning rules of origin; (7) greater integration of developing countries as well as monitoring the WTO/GATT rules to ensure fair treatment of those countries; (8) attention to the problems of anti-dumping rules and their compatibility with some of the Uruguay Round results; and (9) the problem of integrating the 'economies in transition' (e.g. China, Russia, etc.) into the WTO system.

4. DECISION MAKING AND THE WTO: THE 'SOVEREIGNTY' ISSUES

The governing structure of the WTO follows some of the GATT 1947 model, but departs from it substantially. At the top there is a 'Ministerial Conference' which meets not less than every two years. Next there are not one, but four 'Councils.' These include one 'General Council,' which seems to have overall supervising authority, can carry out many of the functions of the Ministerial Conference (MC) between MC sessions, and presumably meets at least as often as the GATT 1947 Council (monthly with exceptions). In addition, however, there is a Council for each of the Annex 1 agreements, namely, a Council for Trade in Goods, a Council for Trade in Services, and a Council for Trade-Related Aspects of Intellectual Property Rights.

There is also a 'Dispute Settlement Body — DSB' to supervise and implement the dispute settlement rules in Annex 2. The WTO Charter provides that the 'General Council shall convene as appropriate to discharge the responsibilities of the DSB.' In the same manner there is a TPR Body for the TPRM.

There have been some allegations that the WTO Charter is an important intrusion on 'national sovereignty.' Apart from the general problems of how to define 'sovereignty' in a world that is increasingly interdependent, the WTO contains an elaborate matrix of decision-making procedures bounded by important constraints. Basically, there are five different techniques for making decisions or formulating new or amended rules of trade policy in the WTO: amendments to the agreements, decisions on various matters, 'interpretations,' waivers, and finally the negotiation and implementation of new agreements. (See Appendix B.)

A careful examination of the WTO Charter suggests that apart from the addition of many new subjects to the substantive annexes, the WTO has no more real power than that which existed for the GATT under the previous agreements. This may seem surprising, but in fact the GATT treaty text contained language that was quite ambiguous, and could have been misused (but fortunately was not) to provide rather extensive powers. For example, in Article 25 of the GATT the Contracting Parties acting by majority vote were given the authority to take joint action 'with a view to facilitating the operation and furthering the objectives of this agreement.' This is very broad and ambiguous language.

Under the WTO Charter, although majority of one-nation-one-vote is the fall-back where consensus cannot be achieved, considerably more attention has been given to the question of decision-making in a number of different contexts, and certain restraints have been added, such as increasing the voting requirements for certain actions (to three-quarters of the members for many waivers and for formal interpretations), and a provision in the amending clauses that a country will not be bound by an amendment which it opposes if the amendment would 'alter the rights and obligations of the members.' Likewise, the waiver authority is more constrained and will be harder to abuse. Furthermore, formal 'interpretations,' 'shall not be used in the manner that would undermine the amendment provisions.' Thus there are more legal grounds than under GATT 1947 to challenge any potential over-reaching on the part of the trade system's institutions. The protections for national sovereignty built into the WTO Charter rules on decision-making are substantially enhanced.

Regarding the practice of 'consensus', as established for several decades in the GATT, several characteristics are worth noting. In the GATT, there is no explicit indication of a 'consensus practice', and the word 'consensus' is not used. The reason that the consensus practice developed was partly the uneasiness of governments about the loose wording of GATT decision-making powers, particularly that in GATT Article XXV. Partly because of this uneasiness, the practice developed of avoiding strict voting. Instead, the Contracting Parties have for several decades taken virtually all of their decisions by 'consensus'. Even when a formal vote was required (such as for a waiver), there would generally be a negotiation for a consensus draft text before such text was submitted to capitals for the formal vote.

In the practice of GATT, however, the word 'consensus' was not defined. In the legal sense, if some sort of 'consensus' could not be achieved, the fallback was the loose voting authority of the GATT. In the WTO Charter, however, consensus is defined (at least for some purposes) as the situation when a decision occurs and 'no member, present at the meeting when the decision is taken, formally objects to the proposed decision.' It should be noted that this is not the same as unanimity, since consensus is defeated only by a formal objection by a member present at the meeting. Thus, those absent do not prevent a consensus, nor does an abstention prevent a consensus. Furthermore, the practice in GATT and surely also in WTO is that some countries which have difficulty with a particular decision, will nevertheless remain silent out of deference to countries with a substantially higher stake in the pragmatic economic consequences of a decision. Thus the consensus practice itself involves some deference to economic power. This has certainly been the practice in the GATT, and the WTO Charter provides that the WTO shall be guided by such 'customary practices'.

The WTO is considerably more explicit about the situation where consensus fails. In a few instances, a decision must be by consensus and there is no fall-back to a majority vote. (For example, adding plurilateral agreements to Annex 4, Article X:9: and amendments to the dispute settlement procedures in Annex 2.) In many other situations, when consensus fails there is an explicit fall-back vote, such as three-fourths of the membership. It is considered quite difficult to achieve such a heavy fall-back vote as three-quarters of the membership (not three-quarters of those voting), since often twenty five per cent of the membership is not involved in a particular decision and may not show up at the meeting. Thus the protections of national sovereignty built into the WTO Charter rules on decision-making are substantially enhanced over that of the GATT.

The amending authority (Article X) is itself quite intricate and ingenious. It obviously has been carefully tailored to the needs of the participating nations related to each of the major multilateral agreements (GATT, GATS (Services), and intellectual property). Amendments for some parts require unanimity. Other parts require two-thirds (after procedures in the Ministerial Conference and Councils seeking consensus for amendment proposals). In almost all cases, as mentioned above, when an amendment would 'alter the rights and obligations', a member who refuses to accept the amendment is not bound by it. In such a case,

however, there is an ingenious procedure (partly following the model in GATT) whereby the Ministerial Conference can by three-quarters vote of the members require a holdout member to accept the amendment, or withdraw from the agreement, or remain a member with explicit consent of the Ministerial Conference (i.e. grant that member a waiver). It is therefore very hard to conceive of the amending provisions being used in any way to force a major trading country such as the United States, the European Union or Japan, to accept altered rights or obligations. As stated above, the spirit and practice of GATT has always been to try to accommodate, through consensus negotiation procedures, the views of as many countries as possible, but certainly to give weight to views of countries who have power in the trading system. This is not likely to change.

There still exist some risks in the voting system and practice under the WTO could be crucial in defining and constraining these risks. For example, bloc voting could develop and there have been hints that the European Union with its number of votes (equal to the number of its members who are also Members of the WTO) and the votes of members of a series of association and affiliation agreements (all totalling more than a majority of GATT and, now, WTO members) could be tempted to use this voting strength in a way to achieve some of its trade policy goals (such as a waiver or selection of officers.)

5. DISPUTE SETTLEMENT PROCEDURES

One of the many achievements of the GATT, despite its 'birth defects,' has been the development of a reasonably sophisticated dispute settlement process. The original GATT treaty contained very little on this, although it did specifically provide (in Articles 22 and 23) for consultation, and then submittal of issues to the GATT Contracting Parties. As time went on, however, the practice began to evolve more towards a 'rule oriented' system. For example, in the late 1950s the practice introduced a 'panel' of individuals to make determinations and findings and recommend them to the Contracting Parties. Before that, disputes had been considered in much broader working parties comprised of representatives of governments.

During the next several decades, the Contracting Parties utilised the panel process more and more. Increasingly, the reports began to focus on more precise and concrete questions of 'violations' of treaty obligations. At the end of the Tokyo Round in 1979, the GATT Contracting Parties adopted an understanding on dispute settlement which embraced some of these concepts, and embodied the practice concerning dispute settlement procedures which had developed during the previous decades.

In the 1980s, the dispute settlement panels were for the first time assisted by a new legal section of the GATT Secretariat. The panels began to write reports that settlement procedures include a segment dealing explicitly with the question of responses available to a complaining state, when a defending member of WTO does not conform to its obligations after a dispute settlement procedure. Of course, a nation can still refuse to conform. For example, this could mean that despite a ruling against the United States on a complaint brought by the United States, the President might be tempted to take action that he is authorised to take under Section 301. Such action could be in violation of international obligations. The mere existence of the possible hypothetical authority to take such action would not necessarily be a violation.

A critical provision in the Uruguay Round text concerning dispute settlement in this regard is Article 23. Article 23.1 requires members of the WTO to use the dispute settlement procedures whenever they seek redress of a violation of obligations or other nullification or impairment with respect to matters covered by the WTO Agreement. It also states that members shall 'abide by, the rules and procedures of this understanding.' Thus, several instances where the United States in the 1980s took unilateral and independent action without proceeding through the GATT, would be inconsistent with the new rules. However, they were also inconsistent with the old rules, to the extent that actual trade restraining measures were applied at the border which violated the GATT (such as an increase in bound tariffs).

8. REFLECTIONS AND PERSPECTIVES: FUTURE PROSPECTS

A safe answer to the question asked in the title of this article, is 'both.' That is to say, the new WTO and the results of the Uruguay Round are both a modest step forward with regard to the institutional structure, and a watershed for the international economic system. The modest step can be seen in the WTO Charter as described above. The watershed, however, is the mere fact of creating a definitive institutional arrangement, combined with the extraordinary expanse of the Uruguay Round substantive negotiations. Clearly this marks a considerable departure from the hesitancy, tentativeness, and 'provisional' nature of the GATT system. It puts in place a structure that will necessarily have to face an increasing number of problems during the next few decades.

Almost any reflective consideration of the rapidly increasing internationa economic interdependence leads to recommendations for increasing internationa cooperation in order to cope with the problems of interdependence. No longer can nations effectively implement their national economic regulations because of international constraints and implications. However, the more one turns to international cooperation institutions, the more it begins to be clear that attention must be paid to the 'constitution' of these international institutions, just as extensive attention is given to national level 'constitutions'.

This raises issues about 'governance', such as preventing abuse of power, effectively channelling important information to decision makers, giving constituencies the opportunity to be heard and to have weight of influence on the decision-making processes. These concepts inevitably lead to questions about the appropriate distribution of power over economic affairs in the world, and the degree to which power should be located in an international institution, or a national/federal institution, or in sub-federal or even very local governmental units. The word 'subsidiarity' is sometimes used to describe these general concepts of distribution of power, and the opposing goals of keeping government decisions at levels as close as possible to affected constituents while empowering higher levels of government to take measures that affect broader segments of society.

The WTO as an institution continues to require attention in several respects. First, an important question is whether the new dispute settlement procedures will work effectively. There will indeed be temptations of some member nations, probably the largest, to ignore or undermine the results of dispute settlement procedures when those procedures do not entirely suit their interests. The dispute settlement procedure will lose credibility, and thus will fail in its primary purpose of establishing and maintaining a creditable 'rule oriented' system.

Second, the decision-making and voting procedures of the WTO, although much improved over the GATT, still leave much to be desired. It is not clear how the consensus practice will proceed, particularly given the large number of members. It may be necessary at some point to develop certain practices about voting to constrain misuse of the various voting rules.

Third, with a likelihood of well over 120 nation members, the WTO must soon face up to its internal procedures for effective governance and administration. This suggests the need for some sort of steering group to guide and advise the Director-General, and other officers of the WTO, in developing priorities for agenda and secretariat work, as well as initiatives to meet new problems in the world. The question of such a steering group has been controversial in the past; every country that suspects it would not be a member of such a group, tends to oppose its establishment. However, without such a group informal mechanisms arise that may be even more exclusive and biased to certain types of economic structures than would otherwise be desired by the membership. For example, the 'quad' group consisting of the US, the EU, Japan and Canada, has been very influential of the GATT system. Surely that influence will continue, but the question is will those four governments be joined by a broader group of representative governments while still maintaining a number small enough to be an effective guiding mechanism for the WTO.

Fourth, an important institutional problem will be how to integrate into the WTO new or emerging subjects which arise in the future, such as the problem of environment and trade rules. The amending rules of the WTO may (as they did in

GATT) turn out to be too rigid and difficult to fulfill. If that is the case, how w new subjects be integrated into the system? As mentioned above, it may be th new 'plurilateral agreements' under Annex 4 will be a major device for this, b the more that occurs, the more we will go back to the old GATT difficulty of 'a carte' choices.

Fifth, there is the broad and important issue of enhancing the publ understanding of the WTO and its work. This leads to the subject 'transparency', commonly understood to mean adequate information a openness of the procedures, advocacy, meetings, etc., so that the media a scholarly endeavours can assist in informing the world public about the operati of the WTO.

Sixth, there has already been much comment about the 'agenda' after t Uruguay Round, for the World Trading System. One question often raised whether there should be another 'round', or whether the procedures for pursui future subjects should try to avoid rounds which tend to be complicated a cumbersome. The arguments in favour of rounds, however, may still preva namely that a larger round gives more opportunity for trade-offs betwe different subjects, and raises the issues to the highest level of natio governments where definitive decisions can usually be made, (whereas lo level officials and technicians would not be able to make such compromises)

Many participants will have great influence on the way the WTO responds those issues, but the most significant influence will likely be that of the n Director-General of the WTO. His guidance and diplomacy will shape the ea years of the Organisation in ways that will have a profound impact on the long term shape of its 'constitution.'

Without necessarily providing a complete inventory of substantive subje issues to be faced in the future by the organisation (in addition to the so-call 'left over issues' mentioned in Section 3.*c* above), the following seem to appropriately included on most lists:

(i) A series of subjects, already newly facing the organisation:

- Competition policy in relation to trade rules.
- Non-market economies, state trading, and economies in transition and th relation to the trading rules (e.g., China, Russia, etc.).
- Product (and service) standards and the use of science, raising the question harmonisation or other techniques for facilitating trade flows.
- Cultural, social policy, and structural impediments to trade.

(ii) There are also a series of subjects called 'link issues', which are subje that are often separately considered, but which have important interlinkages w trade. These include:

- Environmental rules and trade policy.

- Labour standards and rules and trade policy.
- Human rights and trade policy (including economic sanctions).
- Monetary policy and its relation to trade policy.
- Arms control and non-proliferation issues and their relation to trade policy.

(iii) Apart from these subjects for a post-Uruguay Round agenda, there are certain basic GATT concepts that need to be reconsidered and possibly altered so that they will be more effective in the future economically interdependent world. These include:

- The most-favoured-nation clause (and problems of the free rider).
- Regionalism and its relation to the multilateral system.
- Reciprocity and whether this concept can effectively handle the problems of rule making.
- National treatment and the need for minimum standards that go beyond non-discrimination or 'equal' treatment.
- Anti-dumping rules (as previously mentioned) and the question of whether these rules are in reality 'safeguards' or 'escape clause' measures, rather than unfair trade measures.

In short, we have put in place an 'evolutionary watershed' structure of institution and rules to round out the international economic system, but any close look quickly reveals that there is much more to do.

APPENDIX A

Annexes to the WTO

List of Annexes

ANNEX 1

ANNEX 1A: Multilateral Agreements to Trade in Goods
- General Agreement on Tariffs and Trade 1994
- Agreement on Agriculture
- Agreement on the Application of Sanitary & Phytosanitary Measures
- Agreement on Textiles and Clothing
- Agreement on Technical Barriers to Trade
- Agreement on Trade-Related Investment Measures
- Agreement on Implementation of Article VI of the General Agreement on Tariffs and Trade 1994
- Agreement on Implementation of Article VII of the General Agreement on Tariffs and Trade 1994
- Agreement on Preshipment Inspection
- Agreement on Rules of Origin
- Agreement on Import Licensing Procedures
- Agreement on Subsidies and Countervailing Measures
- Agreement on Safeguards

ANNEX 1B: General Agreement on Trade in Services and Annexes

ANNEX 1C: Agreement on Trade-Related Aspects of Intellectual Property Rights

ANNEX 2

Understanding on Rules and Procedures Governing the Settlement of Disputes

ANNEX 3

Trade Policy Review Mechanism

ANNEX 4

Plurilateral Trade Agreements
- Agreement on Trade in Civil Aircraft
- Agreement on Government Procurement
- International Dairy Agreement
- International Bovine Meat Agreement

APPENDIX B

Decision Making Authority in the WTO

WTO Decision Making

ORDINARY DECISIONS: ARTICLE IX:1

The WTO Charter specifies that the GATT practice of decision making by consensus is to be continued for the WTO Ministerial Conference (MC) and General Council (GC), but when consensus cannot be achieved, decisions will be made on the basis of the majority of votes cast, with each member having one vote, unless otherwise provided.

INTERPRETATIONS: ARTICLE IX:2

The MC and GC have exclusive authority to adopt interpretations of the WTO Charter and the Annex 1 Multilateral Trade Agreements. Interpretations of the Annex 1 agreements are to be based on a recommendation of the Council for that agreement and require an affirmative vote from three-quarters of the overall WTO membership. Interpretations are not to be used to undermine Article X's amendment procedures.

WAIVERS: ARTICLE IX:3

The MC may waive an obligation under the WTO Charter and the Annex 1 Multilateral Trade Agreements. If consensus cannot be reached, the grant of a waiver requires an affirmative vote of three-quarters of the overall WTO membership. In the case of Annex 1 agreements, the waiver request is to be submitted to the relevant Council (e.g., Goods, Services or TRIPS), which will submit a report to the MC. Article IX:4 provides that any waiver granted shall specify the exceptional circumstances justifying it and a termination date. Waivers are subject to annual MC review.

AMENDMENTS: ARTICLE X

Amendments to the WTO Charter and the Annex 1 Multilateral Trade Agreements may be proposed to the MC by members or Councils. If consensus cannot be reached, the approval of an amendment for submission to the members requires a two-thirds vote of the overall WTO membership. Generally, amendments come into force on acceptance by two-thirds of the members. If it has been determined that the amendment will not affect member rights and obligations, it comes into force for all members at that time. Otherwise, amendments come into force only for those members accepting them, unless by a three-quarters vote of the overall WTO membership, it is decided that if a member does not accept the amendment it shall be free to withdraw or remain a member with permission of the MC. (In the case of GATS, amendments to certain provisions come into force on a two-thirds vote. For the other provisions, they come into effect for those approving them, but are subject to the same three-quarters vote procedure.)

Amendments to certain rules take effect only upon acceptance by all members: WTO decision making and amendment rules (Arts. IX, X); GATT Arts. I (MFN) & II (Tariff Schedules); GATS Art. II:1 (MFN); TRIPS Art. 4 (MFN).

Amendments to Annexes 2 (Dispute Settlement) & 3 (TPRM) can be made by MC action alone (without member acceptance), but for Annex 2, MC approval must be by consensus. Art. X:8.

NEGOTIATING OTHER AGREEMENTS: ARTICLE III:2

The WTO is to provide a forum for negotiations on agreements contained in its annexes. It may also provide a forum for other negotiations on trade relations and a framework for the implementation of the results of such negotiations, by decision of the MC.

ANNEX 4: PLURILATERAL AGREEMENTS

Decision making and amendments under these agreements are governed by the rules contained in each agreement. (Arts. IX:5, X:10). The MC may add trade agreements to Annex 4 on the request of the members party to the agreement. Art. X:9.

APPENDIX C

Some References to Works of the Present Author Containing More Detail

Restructuring the GATT System (The Royal Institute for International Affairs, Chatham House, London, January 1990).

The World Trading System: Law & Policy of International Economic Relations (The MIT Press, 1989).

'The General Agreement on Tariffs and Trade in United States Domestic Law,' *Michigan Law Review*, **66** 249 (December 1967).

World Trade and the Law of GATT, 'Treatise on a Legal Analysis of the General Agreement on Tariffs and Trade' (Bobbs-Merrill Company, December 1969).

Legal Problems of International Economic Relations — Cases, co-authored with W.J. Davey and A. Sykes, 3rd. ed. (West Publishing Company, 1995).

Chapter on United States, in Volume 'The Effect of Treaties in Domestic Law', F.G. Jacobs (University of London) and S. Roberts (King's College, London), (eds.), (London, Sweet & Maxwell 1987, proceedings of United Kingdom National Committee of Comparative Law Conference in London).

'Status of Treaties in Domestic Legal Systems: A Policy Analysis', *The American Journal of International Law*, **86**, 2, 310–340 (April 1992).

'World Trade Rules and Environmental Policies: Congruence or Conflict?' *Washington & Lee Law Review*, **49**, 4, 1227–1278 (Fall 1992).

'Regional Trade Blocs and GATT', *The World Economy*, **16**, 2, 121–131 (March 1993).

'Managing the Trading System: The World Trade Organization and the Post-Uruguay Round GATT Agenda' Ch. 3, 131–152 in P.B. Kenen (ed.), *Managing the World Economy: Fifty Years After Bretton Woods* (Institute for International Economics, September 1994).

31

"BEST ORIGINAL AVAILABLE"

The WTO and the Global Economy

Richard Blackhurst

IPPINCOTT LIBRARY
OCUMENT DELIVERY

WHARTON REPROGRAPHICS

1. INTRODUCTION

THE ongoing integration of the world economy ('globalisation') and the World Trade Organisation as it approaches its third birthday, are the subject of a great deal of discussion and press coverage. This article focuses on an assessment of the WTO in the context of the evolving global economy. Section 2 reviews recent developments pertaining to the WTO. Several challenges facing the WTO, now and on the horizon, are briefly reviewed in Section 3. The WTO's ability to deal with these challenges is examined in Section 4. The discussion shifts, in Section 5, to a brief consideration of findings regarding the importance of institutions to *national* economic development, and of their implications for the role of the WTO in *global* economic development.

2. WORLD TRADE ORGANISATION

The current 'state' of the WTO can only be described as 'very good'. Implementation of the Uruguay Round results has progressed largely on schedule.[1] Two major post-Uruguay Round liberalisation agreements have been signed — the Basic Telecoms Agreement and the Singapore Ministerial Declaration on Trade in Information Technology Products (ITA). The negotiations on financial services have received a new lease on life and may be concluded in the near future.

RICHARD BLACKHURST has been Adjunct Professor of Economics at the Graduate Institute of International Studies in Geneva since 1974; a Member of the GATT/WTO Secretariat from 1974 until April 1997; and Director of Economic Research and Analysis for the last twelve years of that period. He would like to thank Sven Arndt, Clem Boonekamp, Chulsu Kim, Jesús Seade and Peter Tulloch for very helpful comments and suggestions on key parts of earlier drafts. Any factual errors, incorrect conclusions or naive recommendations are the author's own responsibility.

[1] One part of the implementation involves commitments regarding notification requirements, which number well over 200 following the Uruguay Round. This led members to establish a Working Group on Notification Obligations and Procedures, with a mandate to arrive at recommendations on means to simplify and standardize notifications, and to improve compliance and transparency. See WTO (1996, pp. 110–24) for details.

© Blackwell Publishers Ltd 1997. 108 Cowley Road, Oxford OX4 1JF, UK
and 350 Main Street, Malden, MA 02148, USA.

WORLD ECONOMY 20 (5): 527 - 544

By all counts, the new dispute settlement procedures are working very well, both in encouraging countries to resolve disputes before they reach the panel stage, and in demonstrating a capacity to overturn policies even in sensitive areas. WTO membership in April 1997 reached 131 (up 40 per cent from 93 when the Uruguay Round was launched in 1986), and 28 working parties have been established to consider applications for membership, including from China, the Russian Federation, Chinese Taipei, Vietnam and Saudi Arabia.

The 'single undertaking' nature of the WTO — in contrast to the GATT, every WTO member must submit schedules of tariff 'concessions' and services commitments, and with only *very* limited exceptions (mainly for the least-developed countries) every member accepts in its entirety a common set of rules and disciplines covering goods, services and intellectual property — together with the accession of a number of middle- and lower-income developing countries, has led to a very active WTO programme of technical assistance. These and many other activities are detailed in WTO (1996, ch. V).

The WTO's first two years were capped by three major successes. The Basic Telecoms Agreement and the ITA Agreement have already been mentioned. The third is the aforementioned Ministerial Conference in Singapore. Above all, the results of the Conference amount to a ringing *political* endorsement of multilateralism by WTO members. Among the specific results was the agreement to establish working groups to examine the full range of issues surrounding the interaction between trade and investment, and between trade and competition policy.[2] Depending on the outcome of the working groups' deliberations, the WTO's mandate could be extended to cover comprehensively national policies in these two areas which, while ostensibly domestic, are in reality increasingly trade-related.

The WTO is on a roll. This is a very welcome development, not the least because it's going to need all the momentum it can muster to deal with a number of current and looming challenges.

3. CHALLENGES FACING THE WTO

This section identifies ten 'challenges' facing the WTO. In an order that runs loosely from specific to more general, the ten challenges are:

---

[2] At Singapore there was also agreement (i) to establish a working group to study the issue of transparency (read 'corruption') in government procurement practices, and (ii) to undertake exploratory work on 'trade facilitation', that is, the simplification of trade procedures (such as customs formalities and licensing procedures) in order to assess the scope for improved and expanded WTO rules in this area.

Regarding the interaction of trade and investment issues, WTO (1996, ch. IV) builds a very strong case for bringing national rules on FDI under new WTO disciplines.

© Blackwell Publishers Ltd 1997

- *Achieving a successful outcome of the negotiations on financial services.* The stakes are high for the financial services sector, for the GATS and for the WTO. If the United States is not willing to make a full MFN commitment, it is very unlikely that countries would be prepared to extend the existing interim agreement, in which case the future of financial services in the GATS would be very much in question. Fortunately, there appears to be a commitment of the private sector in some of the major countries to a successful outcome. And there clearly is scope for improved offers if governments are willing to reflect in their GATS schedules the financial reforms and deregulation which many of them are undertaking.
- *Implementing all of the Uruguay Round results on schedule.* With some high-profile parts of the so-called 'built-in agenda' due for implementation a bit further in the future — including all VERs eliminated by the year 2000, the launching of new negotiations on (at least) agriculture and services by 2000, and the critical final 18 months of the MFA phase-out (mid 2003 through 2004) — it is very important for the WTO's credibility that full implementation of the parts due in the next year or two be achieved on (if not ahead of) schedule. It is less a matter of overcoming political difficulties than of maintaining interest and political momentum in the face of a very heavy WTO work programme that includes more 'glamorous' issues.[3]
- *Maintaining the excellent record of the new dispute settlement procedures compiled in the first two years.* There are some very high-profile cases nearing completion or in the pipeline, including those involving the EC's banana regime, the United States–EC dispute over the use of growth hormones in cattle, the Fuji-Kodak dispute and the environmentally-motivated US import ban on a certain shrimp. The attitude of the United States on the proposed (now withdrawn, at least for the time being) panel on the Helms-Burton dispute — US officials said they would ignore it — was *very* worrying. A failure to play by the rules on dispute settlement would strike at the very foundation of the WTO system. If the dispute settlement system does not work, the WTO does not work.
- *The pending accessions of China and the Russian Federation.* It is always difficult to decide at what point a country in transition from central planning to a market economy is far enough along in the process to permit it to observe the full range of WTO obligations and to otherwise play a constructive role in WTO activities. From an economic rather than

---

[3] In addition to the 'WTO Activities' in WTO (1996, ch. V), details on the WTO work programme can be found in WTO document (WT/L/88, 19.10.95) entitled 'Provisions for Review, Future Work or Negotiations in the WTO Agreement and Related Decisions and Declarations'; in the 1996 annual reports of the WTO bodies compiled in WTO document WT/MIN(96)/2; and in the Declaration issued at the end of the December 1996 Ministerial Conference. All three documents are available from the Secretariat, as well as on the WTO website (http://www.unicc.org/wto).

© Blackwell Publishers Ltd 1997

institutional viewpoint, the situation is further complicated by the role which WTO membership can play in the reform process. While early entry into the WTO can help underpin the reforms and prevent backsliding, an undue easing of the entrance requirements can slow the reform process by undercutting officials pushing the reforms. And because the WTO does not exist in a political vacuum, when a country pursuing membership has the trade and geo-political importance of China or the Russian Federation, the entire process can be even more complicated. The member countries face the difficult task of identifying the right terms and conditions of accession for these economies which will enable them to join the WTO, while at the same time not leading to any weakening of the multilateral trading system.

- *Integrating a large number of least-developed and other low-income developing countries into the WTO system.* The 'single undertaking' nature of WTO membership presents a major challenge to many current WTO members and to a majority of the 28 countries in the process of acceding. Two-thirds of the least-developed countries already in the WTO have no representation in Geneva. For the other one-third, it is generally the case that there is only one person, or at the most a very small number, to cover not just the WTO, but also at least some of the other international organisations in Geneva. What makes this such a serious problem is the fact that the WTO is a 'member driven' organisation, in which the delegates play a much more active and important role in the day-to-day activities than is true of other international economic organisations.[4] Nor is the situation of the low-income countries any better when it comes to back-up support in capitals for the heavy workload in Geneva. Unless a major effort to assist these countries is launched very soon, it borders on pure fantasy to believe that they will be in a position to participate meaningfully in the WTO negotiations on (at least) agriculture and services that must begin before the year 2000. If they have not done so already, the demands for technical assistance will soon exceed by a *very* wide margin the ability of the WTO Secretariat to deliver.
- *Competition from regionalism.* It is reasonable to conclude that for much of the post-war period free-trade areas and customs unions have complemented the multilateral trading system (WTO, 1995a). There is, however, no guarantee that this positive relationship will continue, in part because such agreements can be either outward or inward oriented, and in part because of the many problems surrounding the operation of the existing GATT/WTO rules and procedures governing regional agreements — problems which raise the prospect that the post-war complementarity was the result of luck

---

[4] See Blackhurst (forthcoming). National delegates also play an important role in the OECD's work.

© Blackwell Publishers Ltd 1997

or special circumstances that may not hold in the future. The increasing number of 'hub and spoke' agreements, where the spokes benefit less than the hubs (not to speak of third countries), is not reassuring in this regard. More recent concerns have also focused on the risk that the continuing spread of such agreements could lead to a fragmentation of the multilateral trading system and a shift of political momentum from multilateralism to regionalism.[5] The success of the Singapore Conference, together with the Basic Telecoms and ITA agreements, suggest that these concerns are, at least for now, less pressing. An effective response to the other nine challenges on this list would be the most important way for WTO members to ensure the future primacy of the multilateral system.

- *Dealing with the backlash against globalisation*. In a number of countries there are signs of public reaction against the ongoing integration of the global economy. In the rich countries the rising share of GDP traded internationally and increased outflows of FDI are alleged to cause, among other things, job losses, a widening gap between the wages of unskilled and skilled workers and pressures for major cut-backs in social welfare programmes. Some of these charges are echoed by critics in developing countries, along with the claim that globalisation is marginalising the poorest countries and, more generally, widening the gap between the rich and poor countries.

  There are good responses to all of these charges against globalisation (see for example IMF, 1997). From a WTO perspective, the most important one is that regardless of what is really behind each of these developments, increased protection and resistance to further liberalisation can only make matters worse.

  To a large extent, governments have only themselves to blame for the backlash. Part of the problem is their frequent lack of candour regarding the fact that even though globalisation clearly benefits the country as a whole, some groups in the economy will lose. They leave the door wide open for critics to jump on this fact as though it was a completely unexpected result that tips the balance against globalisation. More important is their reluctance to publicly acknowledge that while globalisation creates tremendous opportunities, it also raises the cost of bad policies — such as labour market rigidities, over-regulation, sub-standard schools and educational policies, tax policies that discourage job creation, and chronic budget deficits that reduce investment by reducing national savings. Most

---

[5] See WTO (1996, p. 141) and Carnegie Endowment for International Peace (1997). In January 1997 the European Commission gave strong support to proposals to clarify and otherwise improve WTO rules and procedures governing free-trade areas and customs unions (European Report, 1997).

© Blackwell Publishers Ltd 1997

do not go so far as to actually join in blaming increased trade and FDI outflows for the various developments described above, but their silence on the real culprits creates a vacuum which critics of globalisation are only too willing to fill. Clearly the WTO needs to participate much more forcefully in this debate — stressing the opportunities and answering the criticisms — to guard against an erosion of public support for continued trade liberalisation and a strong multilateral trading system.

- *Ensuring the WTO keeps pace with changes in the global economy.* As trade expands more rapidly than output, a growing share of domestic output is being traded internationally.[6] Along with large multinational firms, small and medium size enterprises in countries at all stages of development increasingly spread their activities — sourcing, marketing, investment — across national boundaries. The WTO, whose primary function is to provide rules designed to reduce policy-related uncertainty surrounding transactions across national frontiers, needs to reflect these and other modern commercial realities. This will require further moves 'inside the border' to cover policies that ostensibly are domestic but in reality have become part of countries' international economic policies.[7]

  Other considerations are pushing in the same direction. As countries' 'economic lives' become more intertwined, it is increasingly likely that trading partners will perceive 'spillovers' from ostensibly domestic policies, for example, competition policies. Success in reducing many conventional border protection measures is also a factor, causing non-border restrictions on market access to become more visible to trading partners (for example, domestic regulations) and tempting countries to compensate for reduced border protection by introducing new non-border protection. For all these reasons, the working groups agreed to at Singapore — on investment and competition policies, as well as transparency in government procurement — need to get up and running as quickly as possible so they can produce results in time for WTO members to decide whether one or more of these topics will be on the agenda, along with agriculture and services, of the next round of negotiations.

- *Preparing for the next round of multilateral trade negotiations.* Decisions taken at Marrakesh require that another round of trade-liberalising

---

[6] For the world as a whole, the ratio of trade to GDP roughly doubled between 1950 and 1974, from 7 to 15 per cent. In the past two decades it has increased to about $22\frac{1}{2}$ per cent, and there is no reason to believe it will not continue to rise for the foreseeable future (WTO, 1995b, pp. 15–23).

[7] A majority of WTO members accept that what they have given up in trade policy discretion by accepting WTO rules and disciplines is more than compensated by the increased predictability and stability of trade policies (at home as well as abroad). Much the same thinking is behind efforts at international rule-making in new areas such as the treatment of FDI.

© Blackwell Publishers Ltd 1997

multilateral negotiations, covering agriculture and services, begins before the year 2000 (there is some quibbling about the precise date). It would be possible to limit the negotiations to two compartmentalised negotiations, similar to the ones on basic telecoms and financial services. The initial experience with those two services negotiations raised doubts about this approach, but at least some of those doubts have been dispelled by the recent telecom results. A more ambitious option — and one more in line with what is needed to prepare the WTO for the first part of the 21st century — would be to launch a major new round with several items on the agenda. Such an agenda also has the advantage of allowing countries to take politically difficult decisions in certain areas in exchange for progress in other areas that furthers their trading interests. Candidates for an ambitious agenda that goes beyond agriculture and services include a further round of tariff cuts and bindings on industrial products, state trading, government procurement, competition policies, rules on foreign investment and reform of the rules and procedures governing free trade areas and customs unions. With just over two years to go, the process of exploring options and deciding on what kind of negotiation to have needs to begin as soon as possible.

- *Getting the members' thinking geared to the WTO's emerging role as the pre-eminent international economic organization.* For much of its nearly 50 years of existence, the GATT was a relatively small organisation, known primarily to delegates in Geneva and a handful of technocrats in capitals. Very few people had heard of 'GATT' or had any idea what it did. The resources which the member countries were prepared to devote to GATT's activities reflected this perception of a small, efficient, relatively obscure institution with a narrow mandate. In the past decade, the actual (and potential) role of the GATT/WTO has been changing dramatically and it is likely to continue to do so, to the point that the WTO soon will be, if it is not already, the pre-eminent international economic organisation. The principal forces behind this changing role are not hard to find: the continuously rising share of national output that is traded internationally (globalisation); the collapse of central planning and the developing countries' rejection of import substitution in favour of development strategies based on integration into the global economy; the success of the ambitious Uruguay Round, including the expanded mandate of the WTO relative to that of the GATT and the 'single undertaking' nature of the WTO; and the approaching near universal membership.

The members' collective attitude regarding the appropriate level of resources to devote to the 'WTO family' — delegates in Geneva, support staff in capitals, and the Secretariat — reflects a kind of modified 1970s thinking. (All references in this article to 'the WTO' refer to an aggregate of these three categories of WTO resources. The occasional use of the term

© Blackwell Publishers Ltd 1997

'WTO family' is intended as a reminder of this aggregate definition.) They know the mandate has expanded since then, and they have made incremental additions to the resources available to the WTO, as well as one relatively major addition (by GATT standards) when the WTO took over from the GATT.[8] But the member countries seem to resist the notion that any fundamental changes have occurred. Implicitly, this amounts to viewing the kinds of challenges flagged in this section as not fundamentally different from those the GATT faced at the beginning of the Tokyo Round — in other words, there may be more work in 1997, but deep down it's still the old GATT. This thinking is evident in the visceral view of many delegates in Geneva and officials in capitals that *any* non-incremental increase in the resources currently available to the WTO family — and, above all, any non-incremental increase in the current size of the Secretariat — would inevitably destroy the 'lean and mean' nature of the WTO and create in its place a bloated, inefficient and under-employed organisation. There is always a chance, of course, that they are right. Then again, as is explained below, it is very likely they are wrong.

4. THE ABILITY OF THE WTO TO DEAL WITH THE CHALLENGES

This section reviews briefly four factors that will determine the capacity (as distinct from the willingness) of the WTO to deal with the challenges just described: the amount of other work to be done; the WTO's mandate; the WTO's decision-making apparatus; and the resources available to the WTO.

a. Other Ongoing Work

Many important parts of the WTO's work programme are covered by the challenges outlined above, in particular those dealing with continued implementation of the Uruguay Round Agreements, dispute settlement, integrating lower-income developing countries into the WTO, getting the working groups agreed to at Singapore up and running, and preparing for the up-coming negotiations on (at least) agriculture and services. But there are also many other parts of the work programme that will demand the attention and time of the WTO in the next two or three years.[9] An illustrative rather than exhaustive list of such work includes the 26 accession working parties other than those for China and the Russian Federation; the 28 working parties examining new free-trade areas and customs unions; trade policy reviews (averaging 15 a year); the Committee on

---

[8] See Blackhurst (forthcoming).

[9] See footnote 3 for a list of references that provide details on the work programme.

© Blackwell Publishers Ltd 1997

Trade and Environment; the Committee on Balance-of-Payments Restrictions; and activities related to closer cooperation with other international organisations (in particular, the World Bank, the IMF and UNCTAD). Although included in the general category of implementing the Uruguay Round Agreements, it is worth noting that the first reviews of certain Uruguay Round results are scheduled for 1998, including the Agreement on Technical Barriers to Trade and the Agreement on Sanitary and Phytosanitary Measures.

b. The WTO's Mandate[10]

Does the WTO's mandate limit its ability to deal with any of these challenges? The mandate can be defined as including two elements: one is the fundamental goals of the GATT/WTO system, and the other is the set of 'instructions' specifying which sectors of the economy and which policies are to be covered in the pursuit of those goals. The GATT system originally had two fundamental goals or functions:

> *First*, by providing a set of rules and procedures governing countries' trade-related policies, it reduced policy-related uncertainty surrounding the exchange of goods across national frontiers. These multilateral disciplines, together with reductions in trade barriers, promoted trade-related investment at home and abroad and brought the gains that come from increased international specialisation.
>
> *Second*, it provided a forum for dispute settlement, and for negotiations both to strengthen and extend the rules and procedures, and to further liberalise trade-related policies.

The instructions regarding the sectors and policies to be covered in pursuit of these goals were simple: the GATT was concerned with trade in goods, and the policy coverage focused on measures applied at the border, in particular, on tariffs and quantitative restrictions.

Between 1947 and 1994 the mandate was enlarged (1) by expanding the goals to include the regular surveillance of members' policies (the TPRM), and the provision of a forum for discussing existing and emerging policy (a change which lapsed with the demise of the Consultative Group of 18), and (2) by expanding the list of policies covered by the rules to include technical barriers to trade, domestic subsidies and government procurement.

In the Uruguay Round, the fundamental goals were expanded in Article III of the WTO Agreement, which spells out the five functions of the WTO.[11] This was

---

[10] The material from here through the end of Section 4 draws on Blackhurst (forthcoming).

[11] The five functions are, respectively, to administer and implement the multilateral and plurilateral trade agreements which together make up the WTO; act as a forum for multilateral trade

© Blackwell Publishers Ltd 1997

not, however, the most important change in the mandate to come out of the Uruguay Round. Much more important was, first, the expansion of the list of sectors covered by the multilateral rules to include services; and second, the expansion of the list of national policies covered by the multilateral rules, in particular to include national laws for the protection of intellectual property, along with sanitary and phytosanitary regulations, and regulations for pre-shipment inspection.

This perfunctory review of the WTO's mandate is sufficient to establish not only that the ten challenges fall within the mandate, but that the WTO would be seriously derelict if it did not do whatever was necessary to fully meet all of the challenges.

c. *Decision-making in the WTO*

WTO councils and committees are open to all members, and in practice the membership is virtually identical on all the ones considered important. There is nothing in the WTO which corresponds to the IMF's or World Bank's Executive Boards or Interim and Development Committees. The closest the GATT came to such a body was the Consultative Group of Eighteen (CG.18), established on a temporary basis in 1975, made a permanent body in 1979 and effectively disbanded in 1988.

It is widely accepted that there is an urgent need to streamline certain categories of work that ordinarily take place in large meetings (this could be viewed as an eleventh challenge facing the WTO). The core of the problem is the unwieldy nature, in many instances, of meetings with a large number of delegates in the room, and a greater reliance on one or more groups (fora) composed of a relatively small sub-set of WTO members will necessarily be at the heart of any solution. The general function of such a group would be to provide a forum for discussing and debating major management issues, important trade issues and new policy areas. Depending on the nature of the topic, the group could also attempt to reach a consensus that would then be presented to the entire membership for approval.[12]

Two competing models for such a group currently are on the table. One involves a formally structured group similar to the previously mentioned CG.18. Besides its relatively small size (not necessarily limited to 18 members), the

---

negotiations; administer arrangements for the settlement of disputes; review national trade policies; and cooperate with the IMF and the World Bank with a view to achieving greater coherence in global economic policy making. The first three functions were part of GATT's original mandate, and the fourth was provisionally added in 1989 and made permanent under the WTO. Only the fifth function is a completely new addition.

[12] Since there is no prospect of the WTO adopting weighted voting, such a group or groups would not — in contrast to the UN Security Council and the Executive Boards of the IMF and World Bank — actually take decisions.

© Blackwell Publishers Ltd 1997

advantages of such a group include its ability to attract representatives from capitals, and the opportunity it would provide to the many small delegations (well over half the WTO membership) to participate — albeit on a rotating basis — in a high prestige, small inner circle of WTO members. The second option, employed in the Uruguay Round and de facto semi-institutionalised at this point — it was used at the Singapore Ministerial Conference — is often described as the 'concentric circles' model. An inner circle of countries serves as the basic discussion, debate and (when necessary) negotiating group. The results are then provided to a second circle or to all the other members, often with a recommendation that a consensus decision be taken on one or more issues. The most important distinction from the CG.18 model is the presence of some flexibility in the membership of the inner circle. Countries which care strongly about an issue generally can gain membership in the inner circle, allowing custom-designed inner circles that ensure the participation of all the members with a strong stake in the outcome on a particular issue.

It is not clear in which direction efforts to find a better forum for discussion, debate and consensus building are likely to go. What is certain is that pressure for a change is bound to continue to build, as the membership grows from today's 131 to more than 160, and as WTO decision-making is faced with a widening range of increasingly more complex issues.

A related but distinct issue concerns the future workability of the principle of taking decisions by consensus in an organisation of 160 plus members. While there are provisions for voting — only the entry into force of certain amendments requires unanimity (WTO Article X:2) — a concerted move away from the 'consensus rule' could have profound implications for an institution based on contractual obligations.[13]

d. The WTO's Resources

As has already been noted, there are three main categories of resources available to carry out the WTO's mandate: the member country delegations in Geneva, their support staff in national capitals, and the WTO Secretariat.[14]

---

[13] Voting very rarely occurred in the GATT, there being a strong preference among members to operate on the basis of consensus, and the members have expressed a preference for the practice to continue in the WTO (because the GATT Council did not have authority to approve waivers and accessions, voting on those matters — via mail ballots — was necessary). WTO Article IX on decision-making states 'The WTO shall continue the practice of decision-making by consensus followed under GATT 1947. Except as otherwise provided, where a decision cannot be arrived at by consensus, the matter at issue shall be decided by voting'. See Blackhurst (forthcoming) for a summary of the provisions on voting in the WTO.

[14] A fourth category of resources, which traditionally has not been important, but which may become more important in the future, involves cooperation with other inter-governmental organisations and non-governmental organisations (NGOs).

© Blackwell Publishers Ltd 1997

(i) *National delegations in Geneva.* A key distinguishing feature of the GATT and now the WTO, relative to other international organisations, is the previously noted point that delegates from member countries play a much more active role in the WTO's day-to-day activities. They like to stress that the WTO is a 'member-driven' organisation, and such considerations as the number of members with permanent delegations resident in Geneva, the size of those delegations, the extent of the individuals' professional experience with GATT/WTO activities, and the support they receive from capitals are all important for the operation of the WTO.

Approximately 80 per cent of WTO members have one or more full-time representatives resident in Geneva. But in virtually every instance they also have to cover other international organisations headquartered in Geneva (UNCTAD, ILO, WHO, WIPO and so forth), which means that most of the national delegations are stretched very thin when it comes to staff available for WTO activities. It is not simply the heavy schedule of activities, but also the level of specialised knowledge required by the expanded mandate. While officials from capitals participate in certain WTO activities, the workload is carried primarily by the Geneva delegations, with the assistance of the Secretariat staff.

(ii) *Back-up support in capitals.* The amount of back-up resources available to the delegations in their respective capitals is very difficult to estimate, in part because much of the assistance is provided by staff who do not work full-time on WTO matters, and in part because it is not enough to consider only staff at the Trade Ministry. Government officials dealing with foreign relations, agriculture, health and safety standards, customs revenue, banking, insurance, telecommunications, patents and trademarks, competition laws and other trade-related policies will, to varying degrees, provide support to colleagues in Geneva. The extent to which the amount of back-up support from capitals varies among the delegations appears to be roughly similar to the degree of variation in the size of the delegations in Geneva. Small delegations in Geneva are likely to have limited back-up support from capitals, and vice-versa for the big delegations in Geneva.

(iii) *The Secretariat.* One way of putting the resources available to the WTO Secretariat in perspective is to compare the staff and budget with those of other organisations. In a ranking comparing the staff and budget of sixteen international economic organisations, the WTO Secretariat ranks 15th on the basis of staff and 16th in terms of the annual budget.[15] The explanation involves not money, but rather the narrow mandate which the WTO members give to the WTO Secretariat.

Undoubtedly the principal explanation of the member countries' desire to keep the WTO Secretariat relatively small, and to give it only a limited role in

---

[15] Recalling the 'member driven' description of the WTO, a very rough guess would put the number of delegates resident in Geneva and working more or less full-time on WTO activities at around 300. If this figure is added to the 513 person Secretariat staff, there are still nine international economic organisations among the 16 with larger staffs. See Blackhurst (forthcoming).

© Blackwell Publishers Ltd 1997

fulfilling the WTO mandate, is the fact that trade policy is so highly politicised at the national level. Governments fear the political consequences at home of any step in the direction of giving the Secretariat more authority and a bigger role. This does not mean they never take such steps, but they do so very cautiously and only infrequently.

Many of the smaller countries, it should be added, apparently would be prepared to accept a larger and more activist Secretariat because they believe it would help them in protecting and promoting their trading interests. Aside from a possible uneasiness about any step that might enhance the ability of smaller countries to protect and promote their trading interests, the larger countries have two incentives for keeping the Secretariat small. Each member's contribution to the Secretariat's budget is based on its share of world trade, which means that increases in the budget are paid for primarily by the larger traders, while many of the benefits from increased Secretariat activities would go to the smaller traders (more on this below). In addition, given the one-country, one-vote nature of the WTO, the large countries have an incentive to see to it that the delegations (that is, themselves) hold on to the bulk of the authority and influence over the nature and direction of WTO activities.

One approach to improving the WTO's resource base would be to finance a minimum number of resident delegates for every country out of the WTO budget. After all, the salaries and other expenses of the executive directors of the IMF and World Bank are paid out of the respective budgets of those two organisations, even though those executive directors represent governments. This could offer a model for dealing with a serious WTO problem, namely the shortage — in 21 instances, the absence — of an adequate number of delegates resident in Geneva and available for full-time work on WTO activities. If the target minimum was three professionals full-time on WTO activities, one option would be to use the WTO budget to finance additional staff for those member countries whose current delegation falls below this minimum.

An easier to administer option (that differs from the first option more in form than substance) would be to provide *every* WTO member with enough funds from the WTO budget to finance three professionals in Geneva. For each WTO member with at least one professional resident in Geneva, there would be a reduction in the Trade Ministry's personnel budget and an increase in the contributions to the WTO. Of course, countries would be allowed — indeed strongly encouraged — to have more than three professionals in their WTO delegations.

Since contributions to the WTO budget are based on shares of world trade, the larger trading nations would be financing to varying degrees the delegations of the smaller traders. An agreement to finance a minimum number of resident delegates in Geneva in all delegations would provide the major trading countries with an opportunity to demonstrate their commitment to a 'member driven' WTO.

© Blackwell Publishers Ltd 1997

e. *Improving Ongoing Activities*

In addition to carrying out its new responsibilities, the WTO must periodically review its performance in more established lines of activity. The Trade Policy Review Mechanism (TPRM) is an excellent example.

Under the TPR mechanism, the four largest traders (the EU, the United States, Japan and Canada) are reviewed every two years, the next 16 largest traders every four years, and the remaining members ever six years, *with a longer interval envisaged for least-developed countries*. The principal rationale behind this schedule presumably is two-fold: first that the trade policies of the biggest traders matter more for world trade than do the policies of the smaller traders; and second, that undergoing a trade policy review is a burden imposed on a country from the outside, which countries reluctantly accept in exchange for the benefit of having the trade regimes of their trading partners regularly reviewed.

One of the most important benefits of the trade policy reviews that have been carried out thus far is the recognition that the principal beneficiary is the country being reviewed. For the first time, officials have a comprehensive view and understanding of their overall trade regime. Transparency appears, often, for the first time, inconsistencies become apparent, and areas most in need of reform become clear, again often for the first time. Subsequent trade policy reviews provide the government with an independent assessment of the progress and direction of their reforms.

These arguments alone are enough to support the proposal to 'TPR' the lower-income WTO members more frequently than once every six years *or longer*. One option would be to do a full review every four years, with an interim review half-way through the four-year period (as is now done with the four largest traders).[16] The case for more frequent trade policy reviews is further strengthened by two additional considerations. Economic reformers are often fighting an uphill battle against entrenched special interest groups and lobbies. A regular dose of transparency and publicity concerning both the pace of trade policy reform and the obstacles to further reform would provide welcome support to reform-minded officials.

Even more important is the potential for using the WTO, via frequent use of the TPR mechanism, to help reforming countries attract badly needed foreign direct investment. Frequent trade policy reviews would allow the WTO to be an 'agency of restraint' that offers foreign investors reasonable assurance, backed up by regular surveillance, that the trade-related reforms will not be reversed (such assurances are also important in eliciting a positive response from the domestic

---

[16] While avoiding any hint of cross-conditionality, the timing of reviews might be loosely coordinated with the IMF's Article IV consultations, in order to bring out both the role of trade policies to the country's overall economic performance, and especially the importance of macroeconomic support for the structural changes set in motion by trade policy reforms.

© Blackwell Publishers Ltd 1997

rivate sector to the economic reforms). Reforms based on WTO obligations bviously fall in this category. By taking advantage of the right to bind reforms and not just reduced tariffs) in its WTO schedules, a country can use the WTO to rovide a broad guarantee that trade-related economic reforms will not be eversed.

Technical assistance is another WTO activity under resource pressure. The eed of many low-income members for technical assistance on WTO-related ssues exceeds by a wide margin anything seen under the GATT. The WTO is a ore complicated agreement and new members are often poorly equipped to andle its demands. A boost in Secretariat resources will be required to help ower-income members deal with the new system and prepare for the next round f negotiations.[17]

5. THE GLOBAL ECONOMY NEEDS A STRONG WTO

The case for ensuring that the WTO has the resources and mandate appropriate the demands of the emerging global economy receives a strong boost from the ndings of research on the determinants of growth at the national level. High vels of savings and investment, technological advances, giving priority to lucation, and similar characteristics of a society have traditionally dominated sts of growth-promoting factors. Stimulated by the work of Douglass North on merican and European economic growth, and four decades of post-war perience in developing countries, economists are adding the quality of a tion's institutions to the list of factors which must be included in the analysis of onomic growth.

This is, in most respects, a relatively new development. A characteristic which ternational economics textbooks share with textbooks in many other areas of odern economics is a neglect of transactions costs, property rights, institutions d rules of the economic game. Traditionally taken as given, these central pects of economic life are not allowed into the analysis. Sensing their portance for understanding economic growth, and building on the work of humpeter, Coase, Alchian, Cheung, Demsetz and others, North developed a eory of institutions and applied it to economic history, an achievement for hich he was awarded (along with Robert Fogel, another economic historian) the 93 Nobel Prize in Economics.[18] One important line of his analysis is evident in following quotation:

For a proposal, including establishment of WTO regional offices, see Blackhurst (forthcoming).

See Myhrman and Weingast (1994) for a short but informative overview of North's tributions to economics.

lackwell Publishers Ltd 1997

> The institutions necessary to accomplish economic exchange vary in their complexity, from those that solve simple exchange problems to ones that extend across space and time and numerous individuals. The degree of complexity in economic exchange is a function of the level of contracts necessary to undertake exchange in economies of various degrees of specialization. Non-specialization is a form of insurance when the costs and uncertainties of transacting are high. *The greater the specialization and the number and variability of valuable attributes, the more weight must be put on reliable institutions that allow individuals to engage in complex contracting with a minimum of uncertainty about whether the terms of the contract can be realized.* Exchange in modern economies consisting of many variable attributes extending over long periods of time necessitates institutional reliability, which has only gradually emerged in Western economies. There is nothing automatic about the evolution of cooperation from simple forms of contracting and exchange to the complex forms that have characterized the successful economies of modern times (North, 1990, p. 34; emphasis added).

In summing up North's contribution to economic analysis, Myhrman and Weingast (1994, p. 193) observe that he has shown us that:

> The wide variations in growth performance across countries and over time cannot be fully understood without explaining the emergence and non-emergence of certain institutions receptive for growth.

As is evident from the two quotations, North focuses on the role of institutions at the sub-national and national level in explaining differences in growth rates among countries and within the same country over time. He makes frequent references to specialisation and the gains from trade, but always within the context of domestic trade within a single country. However, there is no reason to assume that his findings are relevant only for national institutions. Indeed, it seems entirely plausible that, as global economic integration progresses, North's conclusions about the central role of institutions at the national level become increasingly applicable at the multilateral level. Multilateral commerce requires multilateral institutions. The brilliance of the architects of the Bretton Woods institutions (including the GATT) and the OEEC is evident in their ability to recognise this well before it had been conceptualised and empirically verifie even at the national level.

Drawing on both theory and empirical research, North's books and articl elaborate in great detail exactly *how* institutions help to promote economic grow at the national level. In order to gain insights into the role the WTO plays promoting international commerce and economic growth, Petersmann (1997, p 1–65) analyses the WTO as a model for protecting freedom, non-discriminati

© Blackwell Publishers Ltd

and the rule of law across frontiers. A central theme of his analysis is the 'constitutional' function of international organisations, and in particular of the GATT/WTO. Within this context, he identifies three functions of international organisations, which he describes as similar to those fulfilled by national organisations (institutions): to correct market failures, to correct government failures (trade protectionism being a leading example), and to assist in the supply of international public goods (such as the successful Uruguay Round).[19]

The role of the GATT/WTO in avoiding and correcting government failure in the trade area is analysed from a political science perspective in Goldstein (forthcoming). She begins from the well-known proposition that because the gains from freer trade are spread throughout the society, interest group pressures are biased toward those who lose from freer trade. The over-representation of protectionist interests means that without institutional support, it is difficult for a democracy to sustain a liberal trade policy.

Her goal is to explain how international rules and norms affect governments' interest in, and ability to sustain, a liberal trade policy. Four ways in which international institutions affect domestic politics are identified: they act as agenda setters, tying the hands of domestic policy makers and undercutting the influence of protectionist interest groups; by bundling or linking different issues, they make it easier to put together a majority in favour of freer trade; they can change the status of domestic courts and legal institutions; and they can alter the domestic normative discourse.

What might be called a 'double dynamic' is currently at work. The rising share of national output which is traded internationally, and the rising share of national savings which is invested abroad, give the multilateral institutions an increasing role — along with national institutions — in determining each nation's rate of economic growth. In addition, as globalisation progresses and the nature of international commercial relations becomes more complex, the multilateral organisations which provide the institutional and legal frameworks for these cross-border commercial activities need to evolve in ways that allow them to continue to keep pace with the changing conditions.

Claims that this amounts to a call for creating supra-national institutions and transferring greater or lesser amounts of authority from national capitals to those institutions are wrong, at least as regards the WTO. True, the WTO lays down rules that reduce the amount of discretion countries have over their trade-related policies, and it has a dispute settlement system that interprets (but does *not* determine) those rules. But that is very different from an institution with legislative or policy-setting powers. Setting limits on what governments can do, in order to make their trade-related policies more predictable, is a far cry from telling governments what they can or should do. In other words, the argument that

---

[19] See also Roessler (1993).

© Blackwell Publishers Ltd 1997

giving the WTO more resources and more authority amounts to giving a bigger role in world affairs to faceless unelected bureaucrats misses the mark, given that the WTO is — and will continue to be — a 'member driven' organisation. The 'rules of the road' for international commerce are determined and enforced by government representatives from the member countries.

REFERENCES

Blackhurst, R. (forthcoming), 'The Capacity of the WTO to Fulfil Its Mandate', in A. Krueger (ed.), *The WTO as an International Organization* (University of Chicago Press).

Carnegie Endowment for International Peace (1997), *Reflections on Regionalism* (Brookings Institution Press: Washington, DC).

European Report (1997), 'Trade Policy: Commission Calls for Clear WTO Rules on Free Trade Areas', No. 2190 (15 January), 5–6.

Goldstein, J. (forthcoming), 'International Institutions and Domestic Politics: GATT, WTO and Liberalization of International Trade' in A. Krueger (ed.), *The WTO as an International Organization* (University of Chicago Press).

International Monetary Fund (1997), *World Economic Outlook* (May).

Myhrman, J. and B.R. Weingast (1994), 'Douglass C. North's Contributions to Economics and Economic History', *The Scandinavian Journal of Economics*, 96, 2.

North, D.C. (1990), *Institutes, Institutional Change and Economic Performance* (Cambridge University Press).

Petersmann, E.-U. (1997), *The GATT/WTO Dispute Settlement System: International Law, International Organizations and Dispute Settlement* (Kluwer Law International).

Roessler, F. (1993), 'The Constitutional Function of the Multinational Trade Order' in M. Hilf and E.-U. Petersmann (eds.), *National Constitutions and International Economic Law* (Kluwer: Boston).

World Trade Organisation (1995a), *Regionalism and the World Trading System* (April).

World Trade Organisation (1995b), *International Trade: Trends and Statistics* (November).

World Trade Organisation (1996), *Annual Report 1996*, Volume I.

© Blackwell Publishers Ltd 1997

LIPPI
DOCUMENT

The Political Economy of Reciprocity *

H.W. ARNDT

> "In an important sense the GATT does not reflect the views of twentieth century economists and never has done: the exchange of trade barrier 'concessions' is at best in the realm of political economy rather than straight economics" (Snape 1994).
>
> We have "a somewhat surrealist situation – the world's leading organisation in the field of international trade explaining its policies to the world in terms of voodoo economics rejected by virtually every professional economist in the field" (Hudec 1987).

The object of this paper is to answer two questions:

1) If removal of trade barriers is in each country's own interest, why are trade negotiations traditionally conducted on the principle of reciprocity, exchange of "concessions"?

2) Why, if the political economy case for reciprocity is so compelling, have so many countries, especially in the Asian-Pacific region, in recent years nonetheless embarked on unilateral trade liberalisation?

But first some historical background.

□ Research School of Pacific Studies, Australian National University, Canberra, ACT (Australia).

* For helpful comments on the first draft I am indebted to Kym Anderson, Fred Argy, Andrew Elek, Ross Garnaut, David Henderson, Hal Hill, Charles Kindleberger, David Robertson and Richard Snape. The usual caveat applies.

BNL Quarterly Review, no. 190, September 1994.

32

WHARTON REPROGRAPHICS

Concise history of reciprocity

Reciprocity is not new. The notion that a country "favours" another by admitting that country's products more freely to its own markets is implicit even in the most benign form of reciprocity, the most favoured nation (MFN) principle, and this goes back centuries. "Long before the beginning of the 19th century, equality of treatment in tariff matters as between foreign countries had become a sort of principle of international comity" (Viner 1950, p. 5). "The European States began as early as the 17th century to introduce the most favoured nation clause in their commercial treaties" (Haberler 1936, p. 363).

In its more restrictive, bilateral or conditional, form, reciprocity in trade agreements goes back at least to 1794 when the US Congress authorised President Washington to suspend restrictions on imports from other countries "whenever he was satisfied that such countries had suspended their discriminatory restrictions upon the goods and vessels of the United States" (Larkin 1940, p. 23). Such action was taken by several presidents in the following twenty years. Their object, supported by moderate protectionists, was to induce Congress to reduce US tariffs in return for suspension of discriminatory measures by other countries against US trade. The USA throughout the 19th century used the conditional version of the principle of reciprocity: MFN treatment was accorded only to countries guaranteeing it to the USA. In practice, as Haberler (1936, p. 365) pointed out, conditional MFN meant little more than refusal to grant MFN treatment.

In 1922, in the Ford-McCumber Act which gave the President sweeping powers to use tariffs to "equalise costs" between foreign and domestic goods, the USA turned to unconditional MFN, determined o ensure that no foreign country would grant tariff concessions to a hird country without extending the same favour to US goods, but ecognising that "to be prepared to enforce this principle the US must bide by it" (Larkin 1940, p. 27). Conversely, "if the President should oncede tariff reduction on the goods from one country without demanding some similar trade advantage, then every other country would demand the same thing. His bargaining power would be lost" *bidem*, p. 40).

In Britain, the emergence in the 1870s and 1880s of the "Reciprocity Free Trade Association" and similar organisations demanding "fair" rather than "free" trade marked the beginning of the retreat from free trade which culminated in the adoption of protection in 1932 (Bhagwati-Irwin 1987, p. 112).

In 1934, the principle of non-discrimination embodied in MFN became the cornerstone of US trade policy. The Great Depression had led to a worldwide epidemic of beggar-my-neighbour policies which drastically aggravated the contraction of world trade caused by declining incomes. It hit US exports particularly hard because most other countries were compelled or tempted to use balance of payments deficits with the USA to justify protectionist measures. To relieve the pressures on US exports, Roosevelt's Secretary of State, Cordell Hull, secured the passage of the Reciprocal Trade Agreements Act which empowered the President to enter into trade agreements with foreign countries for the reciprocal reduction of tariffs and other trade restrictions, subject to unconditional MFN treatment (Arndt 1944, p. 81). In the following years, many such agreements were concluded.

"Equal access to trade" through non-discrimination became one of the Western Allies' war aims, proclaimed by Roosevelt and Churchill in the Atlantic Charter of 1941 and the basic principle of the Havana Charter of 1946, from which emerged, as the residual legatee of the stillborn International Trade Organisation, the General Agreement on Tariff and Trade (GATT). MFN, however qualified by various exceptions, such as actual or nascent customs unions and preferential arrangements in force in former colonial empires, became the foundation of the GATT code for freer trade.

Reciprocity, appears in the preamble to the General Agreement where the contracting parties express their desire to contribute to its objective "by entering into reciprocal and mutually advantageous arrangements" and in subsequent articles, such as Article 28 which enjoins the contracting parties to "endeavour to maintain a general level of reciprocal and mutually advantageous concessions not less favourable to trade than that provided for in the present Agreement" (United Nations 1947, p. 61).

Reciprocity, as envisaged in the GATT which focuses on matching reductions, "first-difference reciprocity", as it has been called (Bhagwati-Irwin 1987, p. 117), could be seen as helping governments gain acceptance by domestic interest groups for a policy of trade liberalisation and thus for the goal of a free multilateral world trading system. This is much less plausible for the "full reciprocity" version, exemplified more recently by aspects of US and EC trade policy, which aims at broadly matching levels of restriction or openness of trade, both with respect to individual sectors (*e.g.* semiconductors) and bilateral trade balance (*e.g.* US-Japan). This is discriminatory and generally not conducive to free multilateral trade.

Reciprocity is of the essence of international trade negotiations. Freer trade is a means to improved market access. Countries "favour" one another by reducing tariffs. Trade agreements are about reciprocal "concessions". As Snape (1994, p. 7) has put it, "the foundations of GATT that reductions in trade barriers are beneficial, that tariffs are to be preferred to quantitative or other restrictions, that rade barriers should not discriminate between foreign countries and hat trade barriers should be transparent – are fairly sound from an economic perspective. But the concessions framework has many costs: t is often held that it reinforces mercantilist views of international rade. It encourages the idea that other countries' trade barriers are a greater burden than one's own barriers, or even that one's own barriers are not a cost at all".

Why reciprocity?

What accounts for the traditional and almost universal aderence to reciprocity as the guiding principle of trade policy? The answer would seem to be twofold: The superior power of producer iterests concerned to gain or maintain market access in industries ubject to foreign competition and the popularity of "voodoo economics" (in Hudec's sense).

Market access as major policy objective. Tariffs and other trade arriers benefit producers (employers and workers) in protected industries at the expense of (*a*) their foreign competitors, (*b*) producers the country's export industries and (*c*) consumers of their products nd, in so far as protection impairs the efficient use of the country's sources, the public at large). In shaping the country's trade policy, e interests of producers in domestic protected industries will obviously prevail over those of their foreign competitors. They will also

tend to prevail over those of domestic household consumers because producers are more concentrated and better organised. The gains from trade liberalisation have the characteristics of a public good. (The interests of producers in protected import-competing industries may, however, come up against the interests of producers using imported inputs.)

It is primarily to the balance of power between producers in import-competing industries and actual or potential export industries that reciprocity is relevant. Producers in protected import-competing industries may be willing to support "concessions" in the form of lower tariffs if they themselves expect to be able to compensate for loss of market share in the domestic market by gains of market share in exports to foreign markets made more accessible by foreign "concessions". Alternatively, a reciprocal exchange of concessions in a trade agreement may be politically feasible because the interests of producers in protected import-competing industries are overpowered by producers in different (*e.g.* mining or rural) industries.

To put the matter so starkly is to oversimplify. Consumer interests may carry political weight. Diplomatic considerations may favour unilateral concessions to foreign producers, as in the case of preferences to developing countries (perhaps misguidedly – it might well have been in the developing countries' own interest, as Frank 1978, p. 23, has pointed out, to be required to participate in reciprocal tariff reductions under GATT). And of course, governments may be influenced in their trade policies by economists and other advocates of freer trade as in the national interest. But the dominant influence on trade policy has generally been the pursuit of market share by producer interests. National governments have made the enhancement or defence of their industries' market shares the primary objective of trade policy, with reciprocity as the primary negotiating tactic. In Haberler's (1936, p. 376) words, "even countries which are in principle Free Trade must face the *tactical* problem of inducing their Protectionist neighbours to lower their duties". From President Washington's use of reciprocal concessions to diminish discrimination against US exports, to the infinitely painstaking haggling in the GATT Uruguay Round, the primacy of producer interests over those of consumers (or, for that matter, over the efficiency of the economy) has made reciprocity, for most of modern history, the tortuous path to freer trade. Indeed, it has sometimes led governments to impose or raise tariffs as bargaining counters, the so-called "bargaining tariffs" – only to find them hard to bring down again once vested interests have become ensconced. There have also been times, as during the early 1930s, when reciprocity has deteriorated into trade wars, reciprocal infliction of damage rather than reciprocal concessions. Reciprocal trade liberalisation and beggar-my-neighbour trade war are both likely to end up with an unchanged trade balance, but liberalisation at a higher and trade war at a lower level of trade and welfare.

Voodoo economics. Producer interests in protected industries would not find it so easy to insist on reciprocal concessions for any reduction in tariffs were it not for the prevalence of protectionist fallacies among the general public in most countries.

Mercantilist notions – "exports are good, imports are bad" – are deeply ingrained in the minds of most of the unsophisticated public. They readily accept the terminology of a "favourable" balance of trade and of tariff cuts as concessions, not because they still hold to the 17th century mercantilist idea that it is desirable for a country to accumulate "treasure", but chiefly because they think that exports create and imports destroy jobs. They do not make the intellectual effort to visualise that if every country resorts to attempts to export its unemployment, all are worse off. Naive chauvinism ("buy Australian"), obviously vulnerable to the same *reductio ad absurdum,* also plays a part.

Economists concede that protection (and therefore insistence on reciprocity) may sometimes be in a country's national interest. There is the infant industry argument and the optimum tariff argument. The former has had qualified support by economists from Marshall onwards, but its supporters would, in the nature of the case, not be interested in reciprocal concessions to exports. The optimum tariff argument is prominent in US literature on reciprocity. In so far as governments pursue the national interest in tariff bargaining, they are said to "see potential improvements in the terms of trade as the sole motive for departing from multilateral free trade" (Grossman-Helpman 1992, p. 1; also, in more qualified form, Baldwin 1989, p. 126, Hillman-Moser 1992, p. 1). This clearly reflects a "large-country" point of view, a country large enough (or dominant in a particular commodity market) to be able to influence its terms of trade. It is difficult to believe that even US trade policy significantly aims at improvement in the US terms of trade. US trade policy, like that of all other countries, is primarily concerned to improve market access for its exporters and import-competing industries, often at the price of some sacrifice in its terms of trade. While one can think of cases of governments deliberately reducing the volume to improve the terms of trade – the OPEC oil price increases are a notorious example – it is doubtful whether reciprocal concessions are ever sought for optimum tariff reasons.

There are also familiar arguments for temporary protection, to counter predatory dumping or to alleviate balance of payments problems. Neither of these lends support to reciprocity as a principle of trade policy, though both often serve as ammunition for protectionists. (The balance of payment argument for protection – which influenced Keynes in 1932 – needs to be distinguished from the balance of payments argument for reciprocity which, at least in principle, can assist the cause of trade liberalisation: "For balance of payments reasons we can afford to liberalise only if you liberalise, too".)

Clearly, none of these national-interest arguments for protectionist policies rivals voodoo economics as reinforcement of producer interests behind the principle of reciprocity.

To sum up, the conduct of trade policy, nationally and internationally, on the basis of reciprocity, exchange of "concessions", rests on the political power of producer interests out to maintain or increase market share, but derives much of its weight from popular ignorance and prejudice. One should not exaggerate the harm it does in practice. Tariff reductions negotiated on the principle of reciprocity, as in the successive GATT rounds, are better than no reductions. But as the painfully prolonged Uruguay Round has demonstrated, they are cumbersome, slow and hedged around with exceptions and exemptions.

Why unilateral trade liberalisation?

In the past decade, quite a few countries have unilaterally liberalised their trade, cutting tariffs and abolishing import restrictions, of their own accord, without demanding reciprocal concessions from their trade partners. How has this been possible, given the strength of the forces which, as we have seen, have entrenched reciprocity in trade policy for over a century?

Here again, it is worth noting that the recent experience is not without historical precedent. Between 1820 and 1860 Britain moved to free trade, impelled by the confidence of its manufacturers in their competitive superiority (a confidence reinforced by the benefits of cheap food) and persuaded by the free trade advocacy of Adam Smith and Ricardo, Cobden and Bright. By the time Britain concluded the 1860 landmark treaty with France, a treaty formally based on reciprocal concessions, all she needed to offer was to continue her free trade policy (Bastable 1881). France also, under Napoleon III, a convert to free trade, for some years unilaterally liberalised her high tariffs, and there have been similar episodes in the 19th century tariff history of other countries. But it would be difficult to find significant 20th century precedents of across-the-board unilateral trade liberalisation much before the 1980s. (Trade liberalisation within the EC and other preferential trading arrangements has, it need hardly be stressed, taken place within a framework of explicit, though multilateral, reciprocity.)

The forces behind recent unilateral trade liberalisation are best explained by taking two conspicuous examples, Indonesia and Australia. Both countries had for decades (in Australia's case for a century) pursued an inward-looking industrialisation strategy of import substitution in a heavily protected home market, relying on their primary industries for exports. In both countries, it was the slump in commodity prices in the years 1982-85 and consequent balance of payments crisis that triggered the dramatic change of direction from import-substitution to export-orientation, opening up the economy through liberalisation of trade and financial markets. In Indonesia, the collapse of oil prices compelled the realisation that henceforth the country would have to rely increasingly on non-oil exports, *i.e.* manufactures and services. In Australia, the change of outlook which had been under way for some years was highlighted by a warning by the Treasurer (later Prime Minister), Mr Keating, that the country was in danger of becoming a "banana republic" if it allowed the drift of diminishing international competitiveness and mounting foreign debt to continue. In both countries, the political leadership, politicians and technocrats, took the initiative to embark on policies of structural adjustment largely to be effected by exposing domestic industries to international competition. But, as in Britain in

the mid-19th century, the political decisions came against the background of a notable change in the intellectual climate, internationally and nationally.

From around 1970, international organisations published evidence that the import-substitution strategy in relatively small heavily protected domestic markets was a recipe for inefficiency and stagnation (Little *et al.* 1970, Asian Development Bank 1971). Taiwan, South Korea, Hong Kong and Singapore moved towards export-oriented industrial development with spectacular success, and so in lesser degree did some other developing countries. At the same time, faith in socialism, planning and direct controls, gave way the world over to increasing willingness to rely largely on market forces for resource allocation. The decisions to embark on unilateral trade liberalisation in Indonesia, Australia and elsewhere in the 1980s were part of this process.

In Indonesia much of the impetus came from abroad, the World Bank, the IMF, the IGGI (Inter-Governmental Group for Indonesia) consortium of aid donors, and such advisers as the Harvard Institute for International Development, in interaction with Western-trained Indonesian technocrats (*cf.* Hill 1990, MacIntyre 1992). In Australia, economists outside and inside the public service, especially Corden and Rattigan, and one or two politicians, especially Bert Kelly and John Hyde, played a key role, preaching free trade and presenting arguments and evidence (*cf.* Rattigan 1986, Paul Kelly 1992).

If reciprocity in trade policy has largely reflected the pursuit of market shares by import-competing manufacturers and other producers, what interest groups have been behind the unilateral trade liberalisation? In Indonesia, private business has not until recently been identifiable as an important lobby. But there is no doubt that the large financial-commercial-industrial conglomerates (largely Chinese-Indonesian and some related to the President's family) and the Japanese and East Asian joint ventures, with their export interests and international links, now dominate the private sector. In Australia, the mining and rural lobbies which have always favoured free trade have in the past decade been joined not only by more outward-looking financial and other service-industry enterprises but also by manufacturers who have come to see the benefits of moving into foreign markets.

In both Indonesia and Australia high-cost sections of manufacturing industry continue their fight to retain protection, and in

Australia voodoo economics maintains a powerful hold on much of public opinion and parts of the media; indeed, the recent recession saw, as usual, a recrudescence of protectionism and vociferous denunciations of "economic rationalism". The battle for freer trade is never conclusively won. And while protected producer interests try to hold on to what they have – whether US manufacturers lagging in technology or motorcar manufacturers unable to hold their own without protection in Australia or Malaysia, or French and German farmers sheltering behind the Common Agricultural Policy – national trade policies will largely continue in the GATT reciprocity framework, each country jealously trading concessions for concessions. In optimistic moments, one likes to think that, as the benefits of trade liberalisation in cheaper imports and better overall economic performance come to be appreciated, the public's attachment to voodoo economics is weakening. In pessimistic moments, one may grant Bhagwati-Irwin (1987, p. 123) that "proponents of freer trade ... must come to terms with the powerful appeal of the fairness-in-trade doctrine. Without reciprocity, free trade itself would be overwhelmed politically".

But the unilateral adoption of trade liberalisation in so many countries in the Asian-Pacific region has given rise to a new approach that may achieve freer trade without the troublesome handicap of explicit reciprocity. The Asia Pacific Economic Cooperation (APEC) forum aims to encourage and promote unilateral trade liberalisation by individual member countries on a non-discriminatory (MFN) basis (Elek 1992, Garnaut-Drysdale 1994). Each member country is encouraged to liberalise its own trade, knowing that it would reap not only the benefits of freer trade to itself but also improved market access to the markets of other liberalising countries, as has indeed already happened to a considerable extent in the region. Regional economic cooperation for unilateral trade liberalisation would amount to a kind of implicit reciprocity. Whereas explicit reciprocity is liable to be inimical to freer trade since it makes liberalisation by any one country conditional on liberalisation in others which may not happen, implicit reciprocity is wholly beneficial since it takes assurance of improved market access abroad on trust. It is much to be hoped that the APEC experiment will succeed.

REFERENCES

Arndt, H.W. (1944), *The Economic Lessons of the Nineteen-Thirties*, Oxford University Press, London.

Asian Development Bank (1971), *Southeast Asia's Economy in the 1970s*, Longmans, London.

Baldwin, R.E. (1989), "The political economy of trade policy", *Journal of Economic Perspectives* 3(4), Fall.

Bastable, C.F. (1881), *The Commerce of Nations*, 10th ed., 1927, Methuen, London.

Bhagwati, J.N. and D.A. Irwin (1987), "The return of the reciprocitarians – US trade policy today", *The World Economy* 10(2), June.

Elek, A. (1992), "Pacific economic co-operation: policy choices for the 1990s", *Asian-Pacific Economic Literature* 6(1), May.

Frank, I. (1978), "Reciprocity and trade policy of developing countries", *Finance and Development* 15(1).

Garnaut, R. and P. Drysdale, eds. (1994), *Asia Pacific Regionalism: Readings in International Economic Relations*, Harper Educational (Australia), Sydney.

Grossman, G.M. and E. Helpman (1992), "Trade wars and trade talks", mimeo.

Haberler, G. (1936), *The Theory of International Trade*, Hodge & Co., London.

Hill, H. (1990), "Indonesia: export promotion after the oil boom" in C. Milner ed., *Export Promotion Strategies: Theory and Evidence from Developing Countries*, Harvester Wheatsheaf, Hemel Hampstead.

Hillman, A.L. and P. Moser (1992), "Trade liberalisation as politically optimal exchange of market access", mimeo.

Hudec, R.E. (1987), *Developing Countries in the GATT Legal System*, Thames Essay no. 50, Trade Policy Research Centre, London.

Hughes, H., ed. (1988), *Achieving Industrialization in East Asia*, Cambridge University Press, Cambridge.

Kelly, P. (1992), *The End of Certainty: The Story of the 1980s*, Allen & Unwin, Sydney.

Larkin, J.D. (1940), *Trade Agreements: A Study in Democratic Methods*, Columbia University Press, New York.

Little, I.M.D. *et al.* (1970), *Industry and Trade in Some Developing Countries: A Comparative Study*, Oxford University Press, London.

MacIntyre, A. (1992), "Politics and the reorientation of economic policy in Indonesia" in A. MacIntyre and K. Jayasuriya eds., *The Dynamics of Policy Reform in South-East Asia and the South-West Pacific*, Oxford University Press, Singapore.

Naya, S. *et al.*, eds. (1989), *Lessons in Development: A Comparative Study of Asia and Latin America*, International Center for Economic Growth, San Francisco.

Rattigan, A. (1986), *Industry Assistance: The Inside Story*, Melbourne University Press, Melbourne.

Snape, R. (1994), "Economic ideas and trade and development" in R. Garnaut *et al.* eds., *Sustaining the Development Process*, Cambridge University Press, Sydney.

United Nations (1947), *General Agreement on Tariffs and Trade*, New York, vol. I.

Viner, J. (1950), *The Customs Union Issue*, Carnegie Endowment, New York.

33

WHARTON REPROGRAPHICS

International sanctions as international punishment

Kim Richard Nossal

Much of the continuing debate over international sanctions, among scholars and political leaders alike, has focused on their "effectiveness." On the one hand, there is a pervasive view that the many applications of sanctions in the twentieth century[1] show conclusively that they do not "work." As M. S. Daoudi and M. S. Dajani have demonstrated in their exhaustive review of seven decades of literature and public statements on sanctions, the view that these measures are an ineffective tool of statecraft has become almost axiomatic.[2] Yet, as Philip Hanson has noted, policymakers do "not seem to have been deterred by any academic conventional wisdom from applying sanctions."[3] Indeed, practitioners of statecraft seem to hold contradictory views: those who are prone on occasion to embrace sanctions eagerly are just as prone to dismiss their efficacy.[4] It is perhaps not surprising that much

My thanks to Thomas J. Lewis, David A. Baldwin, and the reviewers of *International Organization* for their helpful comments and criticisms.

[1] For the most comprehensive survey of sanctions in the twentieth century, see the survey of 103 "episodes" of sanctions by Gary Clyde Hufbauer and Jeffrey J. Schott, *Economic Sanctions Reconsidered: History and Current Policy* (Washington, D.C.: Institute for International Economics, 1985), especially Appendix C, "Case Abstracts," pp. 107–753.

[2] See M. S. Daoudi and M. S. Dajani, *Economic Sanctions: Ideals and Experience* (London: Routledge & Kegan Paul, 1983). David A. Baldwin, *Economic Statecraft* (Princeton, N.J.: Princeton University Press, 1985) likewise demonstrates the degree to which this skepticism has become the "received wisdom" in the literature on economic sanctions.

[3] Philip Hanson, *Western Economic Statecraft in East–West Relations: Embargoes, Sanctions, Linkage, Economic Warfare and Detente*, Chatham Papers no. 40 (London: Routledge & Kegan Paul, 1988), p. 14.

[4] George Shultz's comment in July 1982 is not untypical: "As a general proposition, I think the use of trade sanctions as an instrument of diplomacy is a bad idea. . . . Our using it here, there and elsewhere to try to affect some other country's behaviour . . . basically has not worked." Quoted in Daoudi and Dajani, *Economic Sanctions*, p. 187; see Appendix II, pp. 178–88, for similar comments from other senior American officials. Likewise, Ronald Reagan and Margaret Thatcher have denied the efficacy of sanctions against South Africa, yet both eagerly embraced sanctions against the Soviet Union.

International Organization 43, 2, Spring 1989

© 1989 by the World Peace Foundation and the Massachusetts Institute of Technology

of the literature on sanctions focuses on a dominant puzzle: "If sanctions do not work, why do states continue to impose them?"[5]

To be sure, there is a revisionist view that challenges the pervasive orthodoxy that sanctions do not work. This view has been advocated most forcefully by David Baldwin, who argues that contrary to the repeated assertions of practitioners and scholars, sanctions *do* "work" and that the real problem is the received wisdom. Using a power analysis perspective derived from the literature on social power, Baldwin argues that when viewed as part of a state's repertoire for the exercise of influence toward others in the international system, economic sanctions are indeed an effective tool of statecraft.[6]

There can be little doubt that Baldwin's work makes a significant contribution to our understanding of sanctions, mainly because he eschews the temptation to measure the "success" or "failure" of sanctions in the kind of simplistic terms favored by analysts such as Gary Hufbauer and Jeffrey Schott, who regard the cardinal criterion for "effectiveness" as a change in the behavior of the state against which the sanctions were imposed.[7] Instead, Baldwin is led to his conclusions by a more complex analytical focus that stresses the importance of the multiple objectives and multiple targets of economic statecraft. But while he comes to more sophisticated conclusions about the "effectiveness" of economic statecraft than many other students of sanctions, Baldwin nonetheless shares with them a common desire to find "rational" purposes for these measures. Indeed, his argument is explicitly based on a model of means-end rationality for explanations of foreign policy.[8] For example, he argues that we should not be content with the claim frequently made that sanctions constitute an "expressive" activity,[9] a "release of internal tension" directed primarily at a domestic audience without other

[5] James M. Lindsay, "Trade Sanctions as Policy Instruments: A Reexamination," *International Studies Quarterly* 30 (June 1986), p. 153.

[6] Baldwin, *Economic Statecraft*.

[7] "In determining whether the episode was 'successful,' we confine our examination to changes in the policies, capabilities or government of the target country." Hufbauer and Schott, *Economic Sanctions Reconsidered*, p. 32. For a comparative assessment of the work of Baldwin and Hufbauer and Schott, see Stefanie Ann Lenway, "Between War and Commerce: Economic Sanctions as a Tool of Statecraft," *International Organization* 42 (Spring 1988), pp. 397–426.

[8] For a statement of these analytical assumptions, see Baldwin, *Economic Statecraft*, especially pp. 15–18.

[9] See Johann Galtung, "Pacificism from a Sociological Point of View," *Journal of Conflict Resolution* 3 (March 1959), pp. 69–71. As I understand Galtung's application of the sociological notions of "expressive" and "instrumental" behavior to foreign policy, the latter is motivated by a desire to resolve a conflict between two parties. By contrast, expressive behavior is designed to serve "the function of tension release from the latent intensity" generated by a conflict (p. 69); the settlement of the conflict is of secondary importance. Galtung uses the distinction in his 1967 discussion of sanctions against Rhodesia, but he shifts his definition considerably: he claims that "the expressive function" of sanctions is to send a "clear signal to everyone that what the receiving nation has done is disapproved of." See Johann Galtung, "On the Effects of International Economic Sanctions, with Examples from the Case of Rhodesia," *World Politics* 19 (April 1967), p. 412. For an application of the distinction between

ends. Rather, he urges that we examine episodes of sanction for their "instrumental" purposes—in other words, for the use of a sanction as an "activity carried out because of the results it is expected to produce rather than for its own sake."[10] For Baldwin, as for most others, sanctions are but one of a number of means to foreign policy ends that can be rationally adduced. Thus, the question for him, as for others, is, What are the useful purposes of this form of statecraft?

In this article, I address this orthodox puzzle, but I focus on one useful purpose of international sanctions that tends to be either overlooked or dismissed outright in the literature on sanctions: the purpose of punishment. I argue that exploring sanctions from the perspective of punishment as a purposive human behavior provides us with useful insights into the utilities of international sanctions. My argument proceeds in three stages. First, by exploring the meaning of the term "sanctions," I seek to establish that sanctions can and should be distinguished from other forms of hurtful statecraft. Sanctions constitute a form of "international punishment," despite the obvious problems of using the notion of punishment in circumstances in which there is no legitimate superordinate authority, and despite the widespread skepticism about the punitive utility of international economic sanctions in the literature. Second, I explore the various purposes of punishment—compulsion, prevention, and retribution—to demonstrate that only some of these purposes are understandable when a model of means-end rationality is used. However, for all of the nonrational elements of punishment, I nonetheless argue that we should see these harms as more than mere forms of "expressive" symbolism embraced for their own sake; rather, we should see them as "instrumental" means to a purposive end.[11] Third, I examine the case of U.S. sanctions imposed after the Soviet invasion of Afghanistan as an illustration of the degree to which punitive intent, particularly retributive intent, can motivate sanctions. I conclude that just as we cannot entirely understand punishment as a purposive human activity solely

---

instrumental and expressive behavior as applied to international economic sanctions, see Peter Wallensteen, "Economic Sanctions: Ten Modern Cases and Three Important Lessons," in Miroslav Nincic and Peter Wallensteen, eds., *Dilemmas of Economic Coercion* (New York: Praeger, 1983), pp. 87–129, especially pp. 98–102.

[10] Baldwin, *Economic Statecraft*, p. 16; see also pp. 97–99. It might be noted that Baldwin's interpretation of what constitutes "instrumental" activity, while it is both more logical and more elegant than Galtung's definition in "Pacificism," is nonetheless considerably at variance with that definition.

[11] This is not, therefore, merely Galtung's "expressive" argument about sanctions restated in different form. Punishment may be an activity that is "expressively gratifying" (in the sense that Galtung used that term in 1959 on p. 71 of "Pacificism"); or (as he was to use the same term in 1967 with a rather different meaning) punishment may serve the "expressive function" of signaling disapproval to others, both domestically and internationally. Rather, I argue below that punishment is always an activity that is purposive, self-conscious, and "instrumental" (as Baldwin uses the term), even though its ends (or useful purposes) cannot always be understood when a model of means-end rationality is used.

by reference to a rational model of a means to a clearly delineated end, so too we cannot entirely understand sanctions as a form of international punishment by an attachment to a rational model of policy behavior. But in the nonrational purposes of punishment we may find the key to the sanctions puzzle, for focusing on punishment explains why sanctions as punishment "work" and therefore why this tool of statecraft continues to be embraced by foreign policymakers.

Sanctions and wrongdoing

One of the impediments to clear discussions about the useful purposes of sanctions is the use of the term "sanction" itself. Definitions tend to be idiosyncratic, often sloppy, and frequently in violation of the minimum requirement that a word be defined in a way that generally conforms to common usage. For some, sanctions are those legal policy instruments that are used to enforce international law.[12] For others, sanctions are the instruments of economic coercion.[13] Some, such as Klaus Knorr, only use the term in a sociological sense: sanctions, whether positive or negative, are simply the means of exercising power.[14] For others still, sanctions are the economic instruments used to achieve "foreign policy goals."[15] It is perhaps hardly surprising that Baldwin notes in exasperation that the term "is used in so many different ways that there is much to be said for avoiding it altogether." However, he ends up using it—albeit reluctantly—if only because "ignoring it is impossible." But by using it, Baldwin merely adds to the definitional confusion: he provides his own definition, which simply equates economic sanctions with economic statecraft writ large.[16]

However, as Baldwin himself recognized in his cogent critique of interdependence,[17] definitions matter. When scholarly inquiry focuses on the motivations for using a set of policy instruments, and particularly on the effectiveness of those instruments, coherence can hardly be expected if the investigators cannot agree about the subject of their inquiry. Thus, my examination of the notion of sanctions is not merely a semantic exercise designed to add yet another idiosyncratic definition to an ever-expanding stock. Rather, it seeks to show that economic sanctions can, and should, be distinguished from other types of economic statecraft (if we accept Baldwin's entirely reasonable definition of economic statecraft as economic instruments used by governments for "influence attempts" in international politics[18]). I argue that two distinguishing characteristics can be identified: first, international sanctions are imposed for acts of wrongdoing; and second, they are punitive in intent.

---

[12] See, for example, Hans Kelsen, *The Law of the United Nations* (New York: Praeger, 1950), pp. 760ff; and the discussion in Daoudi and Dajani, *Economic Sanctions*, pp. 2–9.

[13] See Donald L. Losman, *International Economic Sanctions: The Cases of Cuba, Israel and Rhodesia* (Albuquerque: University of New Mexico Press, 1979), in which embargo and sanctions are used interchangeably throughout the book. See also Richard Stuart Olson, "Economic Coercion in World Politics, with a Focus on North–South Relations," *World Politics* 31 (July 1979), pp. 471–94.

[14] Klaus Knorr, *The Power of Nations: The Political Economy of International Relations* (New York: Basic Books, 1975), chaps. 1 and 6.

[15] This is the definition that underlies the mammoth survey of economic sanctions by Hufbauer and Schott. See *Economic Sanctions Reconsidered*, p. 2.

[16] Baldwin, *Economic Statecraft*, p. 36.

[17] David A. Baldwin, "Interdependence and Power: A Conceptual Analysis," *International Organization* 34 (Autumn 1980), pp. 471–506.

It is important to begin by noting that sanctions are invariably imposed in response to some act. Some students, such as Klaus Knorr, Richard Olson, and Donald Losman,[19] tend to focus on the instruments of economic statecraft employed by governments rather than on what prompts their imposition. Thus, all coercive economic policy instruments—embargoes, boycotts, and indeed any disruption in the "normal" or "customary" levels of economic intercourse between states—are too often simply described as economic "sanctions." Likewise, other students focus on the intended effects of sanctions, excluding, for example, cases in which the goals sought by economic means are "economic" and examining only cases in which the goals sought by employing economic instruments of statecraft are "political."[20] The major problem with such approaches is that they obscure the well-established etymological link between sanctions (both positive and negative) and morally wrong (or right) behavior.[21] They thus distort not only the meaning of the term itself but also—and more important—our assessment of the "effectiveness" of this instrument of statecraft.

On the other hand, Margaret Doxey has argued that the element of wrongdoing is critical to our understanding of sanctions: "It is still possible—and desirable—to preserve the sense of sanctions as penalties linked to real or alleged misconduct."[22] Such usage, it might be added, accords well with the Latin origin of the word, which clearly establishes the relationship between wrongdoing and the hurts imposed on wrongdoers. *Sanctio* was the penalty specified for a transgression of a law or decree and particularly for a violation of a sacredness. The Latin origin thus suggests that a sanction is not simply a penalty imposed for a violation of the rules; it is a punishment for a "wrong" deemed to be immoral, or offensive to the moral conscience, and damaging to the interests of the community as a whole. Punishment thus involves not only a concrete cost (the deprivation of something of value) but also a public expression of the community's moral disapproval of the act.[23] When this is applied to international politics, it could be argued that what prompts one state to invoke "sanctions"—not merely the instruments of economic coercion—is the perception that the target state has violated norms of moral behavior valued by the sender and thus deserves not only concrete penalties but also a public proclamation of the target's impiety.[24] Such usage more clearly reflects the etymology of the word as well as its concern with the moral gravity of the violation and (to the extent that there is a "public realm" in international politics[25]) the essentially "public" nature of the "objectionable" act.[26]

---

[18] Baldwin, *Economic Statecraft*, p. 32.

[19] Knorr, *Power of Nations*, chap. 6; Olson, "Economic Coercion," p. 474; and Losman, *International Economic Sanctions*, chap. 1.

[20] Such an exclusion is not uncommon. See, for example, Lindsay, "Trade Sanctions," p. 155; Hufbauer and Schott, *Economic Sanctions Reconsidered*, p. 2; Miroslav Nincic and Peter Wallensteen, "Economic Coercion and Foreign Policy," in Nincic and Wallensteen, *Dilemmas of Economic Coercion*, p. 3; and James Barber, "Economic Sanctions as a Policy Instrument," *International Affairs* 55 (July 1979), pp. 367–68.

[21] *The Oxford English Dictionary* thus distinguishes between "vindictive" and "remunatory" sanctions, but both kinds are clearly linked to wrongdoing (or "rightdoing"), as can be seen in the first usage of "positive" sanctions appearing in 1692: "The strictest Sanction which any Soveraign Power can give unto its Laws is, when it . . . hath . . . declared, That it will conferr a sufficient share of good Things, or Rewards, for so doing; and of Evils, or Punishments, upon any breach, or neglect of its Commands."

[22] Margaret P. Doxey, *International Sanctions in Contemporary Perspective* (London: Macmillan, 1987), p. 4. Compare her earlier, more restricted, definition in *Economic Sanctions and International Enforcement*, 2d ed. (New York: Oxford University Press, 1980), p. 7.

There is, of course, little agreement on what constitutes a "morally objectionable" act in interstate relations. Indeed, much of the literature on economic sanctions sidesteps the thorny and necessarily normative issue of wrongdoing. For example, Hufbauer and Schott refer to the "objectionable" policies of targets but make little effort to indicate why they think that some "objectionable" behavior warranting economic responses (Japanese trading practices, for example) should not be considered as cases of economic sanctions, while other "objectionable" behavior warranting similar economic responses (Zimbabwean criticism of U.S. foreign policy, for example) should be.[27] Doxey, by contrast, does try to establish the nature of international standards by surveying practices in the League of Nations, the United Nations, and various regional organizations and also outside organizational contexts. Her survey demonstrates the limited extent of widely held standards of behavior and the degree to which the responses of the international community to "wrongdoing" are "highly erratic."[28] Doxey's work also demonstrates the difficulty of generalizing about which acts, under which circumstances, will be seen by states as "morally objectionable" and thus deserving of "sanctions." Not only is it difficult (some would say impossible) to establish clear and universally held standards of "right" and "wrong" in the absence of a superordinate authority to give these terms meaning, but responses to acts that are essentially similar in nature will invariably be highly dependent both on context[29] and on the dictates of realpolitik.[30]

---

[23] Joel Feinberg, *Doing and Deserving: Essays in the Theory of Responsibility* (Princeton, N.J: Princeton University Press, 1970), pp. 96–97, argues that we make such distinctions in domestic law between "penalties" for rule breaking that is not considered evil (overstaying one's time at a parking meter, for example) and "punishments" for acts of wrongdoing that are morally offensive to the community and therefore involve both "hard treatment" and condemnation.

[24] Such statements accusing the target with violation of some sanctity have always been an integral component of international sanctions, whether it be Athenian charges of Megaran impiety in cultivating consecrated ground or American condemnations of Soviet violations of "certain irreducible standards of civilized behavior" in downing a civilian aircraft. See Thucydides, *The Peloponnesian War*, trans. Rex Warner (Harmondsworth, U.K.: Penguin, 1954), p. 118; and Seymour M. Hersh, *"The Target Is Destroyed": What Really Happened to Flight 007 and What America Knew About It* (New York: Vintage, 1986), p. 184.

[25] On this, I follow Hedley Bull's assertions concerning "community" in the contemporary international system; see *The Anarchical Society: A Study of Order in World Politics* (New York: Columbia University Press, 1977), chap. 2.

[26] Galtung also stressed the importance of norms in the imposition of sanctions, arguing that the purpose of economic sanctions was to alter behavior that violated norms held by the sender. See Galtung, "Effects of International Economic Sanctions," p. 380. My purpose, by contrast, is not to focus here on the *intended effects* of sanctions but, rather, on what kind of international behavior provokes them. See also Daoudi and Dajani, *Economic Sanctions*, p. 7.

"Punishment" in international politics

To this point, I have argued that sanctions can be distinguished from other forms of statecraft by the type of acts that provoke their imposition. The second distinctive feature of sanctions is their punitive intent. It is usually argued that the intent, or purpose, of sanctions is to cause harm to the target state in order to achieve a number of foreign policy objectives. James Lindsay's examination of these goals is a useful exemplar of this kind of analysis. He suggests that when states use sanctions, they are seeking to achieve one or more of five broad ends: compliance, subversion, deterrence, international symbolism, or domestic symbolism.[31] If, however, as I have argued, sanctions are policy responses to acts perceived by the sender to be acts of moral wrongdoing, it is difficult to exclude the *punitive* objective of sanctions: in other words, the goal of punishing an act of wrongdoing.

---

[27] Hufbauer and Schott, *Economic Sanctions Reconsidered*, p. 27.

[28] See Doxey, *International Sanctions*, p. 143.

[29] The markedly different international reactions to the destruction by a Soviet fighter of Korean Air Lines flight 007 in September 1983 and the destruction by a U.S. cruiser of Iran Air flight 655 in July 1988 amply demonstrate this. In 1983, the downing of 007, with the loss of 269 lives, was widely characterized in Western rhetoric as an evil and morally repugnant act that was deserving of sanctions. Indeed, many Western leaders called it an act of "murder," with the attendant implications of *mens rea* on the part of those who had participated in the decision. In 1988, by contrast, the downing of 655, with the loss of 290 lives, was widely characterized as a "tragic accident." Few were willing to see this as an act of moral wrongdoing, much less attribute murderous intent to those who had authorized the shootdown; there was no international support for Iran's initial demand that sanctions be imposed against the United States.

[30] The degree to which the perceptions of wrongdoing will often depend on an essentially amoral calculus can be seen in the reactions of the U.S. government to violations of one of the cardinal rules in contemporary international law and practice—the proscription against one state sending its troops uninvited into another state and replacing the legitimate government there by force. This rule has been openly broken on numerous occasions since 1975: by Indonesia in 1975 (East Timor), by Vietnam in 1978 (Kampuchea), by Tanzania in 1979 (Uganda), by the Soviet Union later that year (Afghanistan), by Argentina in 1982 (Falklands), and by the United States in 1983 (Grenada). Judging by the U.S. government responses, however, only some of these acts were perceived to be "wrongful" or morally repugnant.

At first blush, it may seem incongruous to use the term "punishment" in the context of international politics. For both the common and philosophic uses of the word in the English language strongly suggest that punishment tends to be understood primarily in the context of the rule of law and legitimate authority. Thomas Hobbes's definition of punishment establishes this relationship clearly: "an Evill inflicted by publique Authority on him that hath done, or omitted that which is judged by the same Authority to be a Transgression of the Law."[32] The Hobbesian formulation explicitly denies that a hurt administered by other than public authority, or for reasons other than in response to a transgression of law, is punishment; rather, in these (and nine other) instances, the harm must, in Hobbes's view, be termed an act of hostility, not punishment.[33] This formulation endures both in political theory[34] and in the contemporary literature on punishment[35]: the authoritative status of the agent inflicting harm on a transgressor of laws remains integral to how in domestic politics we distinguish between acts of punishment and acts which, without that legitimate authority, we would otherwise term theft, battery, forcible confinement, or murder.

In international politics, of course, there exists neither a public authority to inflict legitimate harm on transgressors nor laws (as that term is understood in a domestic political context) that can be transgressed. Indeed, what Alfred Lindesmith says of punishment in primitive societies might well apply to international politics: where there are no "central governments capable of exercising coercive control over the subgroups which constitute the society, it would not be logical to speak of either crime or . . . punishment."[36] It might therefore be reasonably suggested that the hurts administered by sovereign states on each other for whatever reason cannot be called "punishments" at all; instead, such acts fall more properly under Hobbes's "acts of hostility" committed in a state of war.[37]

---

[31] Lindsay, "Trade Sanctions," pp. 155–56. See also Barber, "Economic Sanctions as a Policy Instrument"; and Nincic and Wallensteen, "Economic Coercion and Foreign Policy," pp. 4–8.

[32] Thomas Hobbes, *Leviathan*, ed. C. B. Macpherson (Harmondsworth, U.K.: Penguin, 1968), chap. 28, p. 353; original italicized.

[33] Ibid., pp. 354–57.

[34] Indeed, for some, the right of an individual to be punished is integral to liberal democratic contract theory. See, for example, Thomas J. Lewis, "Contract Theory and the Right to Be Punished," *American Behavioral Scientist* 28 (November–December 1984), pp. 263–78.

[35] Ted Honderich defines punishment as "an authority's infliction of a penalty on an offender"; see *Punishment: The Supposed Justifications* (Harmondsworth, U.K.: Pelican, 1971), p. 19. See also Sir Walter Moberly, *The Ethics of Punishment* (Hamden, Conn.: Archon, 1968), p. 35.

[36] Alfred Lindesmith, "Punishment," in *International Encyclopedia of the Social Sciences*, vol. 13 (New York: Collier Macmillan, 1968), p. 219.

The literature on international sanctions tends to reflect the essence of the Hobbesian connection between legitimate authority and punishment: there is no legitimate authority in international politics; there is no law; and there can therefore be no punishment, only acts of hostility. The problem is that most authors fail to make the careful semantic distinction that Hobbes himself insisted on, and they proceed to use "punishment" and "harm" interchangeably. Robin Renwick, for example, acknowledges that one of the *effects* of economic sanctions is punitive, but it is clear that he is using the term "punishment" simply to mean "hurtful" measures to achieve the other "purposes" of compliance, subversion, or deterrence, rather than hurtful measures intended to be a punishment.[38] Likewise, both Baldwin and Knorr seem to use the term "punishment" simply to mean the infliction of harm or injury.[39] Thus, while Knorr does suggest that one of the goals of economic power is "symbolizing displeasure and inflicting punishment,"[40] it is clear that he is not referring to the punishment of wrongdoing. While technically not incorrect,[41] such synonymous usage tends to focus our attention on the harm itself and to obscure the reasons why the harm is being administered.[42] More important, using "punishment" merely to mean "harm" obscures almost entirely the concept of harm imposed for wrongdoing.

There are certainly some students of sanctions who recognize that punishment is a particular type of harm. Fredrik Hoffmann, for example, notes *en passant* that there are "morally 'good' and morally 'bad' nations, and it is implied in the very term 'sanction' that this measure is not just any political action; it is intended to be used against nations that deserve 'punishment.'"[43]

---

[37] In *Leviathan*, p. 356, Hobbes asserted that "Harme inflicted upon one that is a declared enemy, fals not under the name of Punishment: Because seeing they were either never subject to the Law, and therefore cannot transgresse it, . . . all the Harmes that can be done them, must be taken as acts of Hostility."

[38] See Robert Renwick, *Economic Sanctions* (Cambridge, Mass.: Center for International Affairs, 1981), pp. 91–92: "The purpose of sanctions was conceived initially as being either preventive or remedial. Their main effect, however, has invariably been *punitive*. There are international circumstances in which it may become necessary to take some punitive action, falling short of the use of force, either to weaken the regime to which sanctions are applied or, by penalizing it for one undesirable action, to try to deter it from further action of that kind."

[39] See, for example, Baldwin, *Economic Statecraft*, pp. 264–65; and Knorr, *Power of Nations*, p. 4.

[40] Knorr, *Power of Nations*, p. 138.

[41] "Pain" and "punishment" are derived from the same root: from the Latin *poena*, the money that was paid in atonement for some wrong, which itself is from the Greek *poinē*, payment or penalty.

[42] It is for this reason that we would not likely describe a victim being beaten up by a mugger as being "punished"; by contrast, it would be more appropriate to use the word "punishment" to describe the actions of a victim who was able to turn on the mugger and beat him.

[43] Fredrik Hoffmann, "The Functions of Economic Sanctions: A Comparative Analysis," *Journal of Peace Research* 4 (April 1967), p. 144.

David Leyton-Brown and Margaret Doxey, by contrast, explicitly add punishment to their list of the purposes of sanctions. Leyton-Brown argues that such punishments serve to define the parameters of acceptable behavior and thus add to international order.[44] Likewise, Doxey notes that foreign policymakers frequently "protect themselves from later accusations of failure by asserting that punishment, not target compliance, is the object of the exercise. In other words, sanctions are presented as measures which make life more difficult for the target by attaching a cost, or price, to its conduct; a kind of fine for international misbehaviour."[45]

But Hoffmann, Leyton-Brown, and Doxey are in a minority. Most others appear to be distinctly uncomfortable with the idea that international sanctions might have punitive purposes, a discomfort that has deeply embedded historical roots. For example, the Royal Institute of International Affairs group that prepared a report on sanctions in 1938, during the interwar period, ironically made no mention of the element of punishment in their discussion of the purposes of sanctions. In one odd passage, they went so far as to deny that the hurtful sanctions clearly specified under Article XVI of the League of Nations Convenant were penalties at all:

> The Covenant . . . is based on the view that the object of sanctions is purely that of preventing the success of aggression. It lays down no penalties for the violation of international law. . . . It merely prescribes the methods which Member States are bound to follow in rendering assistance to the victims of aggression.[46]

Such views may have been prudent in the mid-1930s, when there was a common fear in the Western democracies that the imposition of economic sanctions by the League against Germany, Italy, or Japan would precipitate a general war.[47] But it is difficult to argue that the "methods" set out by those who framed the Convenant were not intended as anything but punishments for wrongdoing. The "League experiment" was explicitly designed to bring the conditions of domestic order to the international system. A set of rules was laid out in the Convenant; transgressors were threatened with the imposition of clearly specified hurtful penalties if they broke the "law"; and the penalties were to be exacted by an institution that was supposed to come as close to being an "international public authority" as possible. The penalties against a Covenant-breaking state—harms legitimized by the international community—had all of the structural properties of punishment as we know it in a domestic context.[48]

---

[44] David Leyton-Brown, "Lessons and Policy Considerations About Economic Sanctions," in David Leyton-Brown, ed., *The Utility of International Economic Sanctions* (London: Croom Helm, 1987), p. 303.

[45] Doxey, *International Sanctions*, p. 92.

[46] Royal Institute of International Affairs (RIIA), *International Sanctions* (London: Oxford University Press for the RIIA, 1938), p. 13. Such a view was not uncommon among practitioners at the time: for example, the Ecuadorian delegate to the League Assembly argued in 1935 that "the sanctions envisaged under the Covenant involve no punitive intention, no element capable of wounding the pride of any nation which may embark upon a war." Quoted in Daoudi and Dajani, *Economic Sanctions*, p. 71.

[47] For a brief but useful discussion, see Renwick, *Economic Sanctions*, chap. 2.

Such discomfort with the idea of the punitive purposes of sanctions is no less evident in the present era, when the attempted legalism of the League has long been abandoned. One indicator of this is the tendency to characterize "punishment as a goal of sanctions" as the vaguely irrational product of emotionalism, authoritarianism, or sadism. For example, it is clear from Knorr's brief discussion of the sanctions imposed against India by the United States in January 1972 that he thinks that punishment is merely the result of "emotionalism": economic sanctions, he suggests, can be used to "gratify the actor's emotional desire for hurting a recalcitrant government." He claims that the American measures resulted "from a sheer and petulant desire to punish and rebuke New Delhi."[49] Johann Galtung, for his part, completely dismisses the punitive element at the outset of his examination of economic sanctions against Rhodesia. He acknowledges that sanctions are intended to achieve "either or both of two purposes: to punish the receivers by depriving them of some value and/or to make the receivers comply with certain norms." But he doubts that sanctions can serve *both* purposes: "Modern penology does not seem to warrant much belief in punishment as a *general* method for making people comply." Thus, those who would use punishment even when the prospects of compliance are slim are summarily written off as "highly punishment-oriented in the sense that punishment has become an automatic and probably also cherished goal in itself." Then, noting offhandedly that "this punishment-oriented attitude is probably widespread, particularly as applied to the international system, and serves to maintain negative sanctions," Galtung puts aside the issue altogether: "We shall disregard the punishment aspect and be interested in sanctions only as a way of making other international actors comply."[50]

The issue of "punishment as a goal in itself" has also been raised by other authors. Sidney Weintraub, for example, suggested that when American sanctions against Cuba "failed to achieve this goal [the overthrow of Castro], . . . the primary aim of policy became one of punishment of the communist regime."[51] Weintraub's surmise is altogether too much for Baldwin, for whom such measures can easily be shown to have rationally instrumental purposes ("lowering the attractiveness of the Cuban example to others, limiting Castro's capabilities, and imposing costs on the Soviet Union").

---

[48] Doxey, *International Sanctions*, chap. 3.

[49] Knorr, *Power of Nations*, p. 138.

[50] See Galtung, "Effects of International Economic Sanctions," pp. 380–81; emphasis in original.

[51] Sidney Weintraub, "Current Theory," in Sidney Weintraub, ed., *Economic Coercion and U.S. Foreign Policy: Implications of Case Studies from the Johnson Administration* (Boulder, Colo.: Westview Press, 1982), p. 10.

"Pure, unadulterated sadism," Baldwin asserts testily, "is rarely a foreign policy goal."[52]

One would be hard-pressed to disagree. But Knorr, Galtung, and Baldwin, it can be argued, confuse several important points about the goals of punishment. First, punishment is much more than simply expressive behavior—a "petulant" venting of an "emotional desire to hurt recalcitrants." While an act of wrongdoing invariably gives rise to such expressive emotions as petulance, anger, and vengeance, the punishment of that act is, as I argue below, a purposive and instrumental behavior (and, it might be noted, is meted out in the vast majority of instances without anger, petulance, or other "emotionalism"). In other words, the purpose of punishing a wrongdoer is not simply to vent emotions, although that may be a consequence of the punishment.

Second, punishment cannot, by definition, be a "goal in itself." To say that "one punishes in order to punish" or that "one punishes for the sake of punishing" is to make tautological nonsense of the normal understanding of the word. Punishment must by definition be a purposive behavior, a hurt inflicted for identifiable reasons and with identifiable objectives. If I "punish" someone, it is by definition "for something"—in two senses. First, the person punished must have done (or be thought to have done) something "wrong." Second, the act of punishment is by definition an act designed to achieve an end or ends related to the wrongful act (with the end usually but not always being the prevention of its recurrence). By contrast, the same is not true if I "hurt" someone. I may hurt without the target having done anything "wrong" to provoke my harm; likewise, I may hurt capriciously, with no particular objective or goal in inflicting the harm. In short, it is the purposiveness of the action that gives the word "punish" its meaning.

Third, Baldwin asserts and Galtung implies that punishment and sadism are basically the same. But simply equating the two muddies distinctions that are important to maintain. Sadism has but one goal: the creation of pleasure by inflicting hurt and pain on another. Sadists need not predicate their infliction of pain on the prior behavior of their targets, and they will be indifferent about the effects of their attacks on the subsequent behavior of their targets or of others. And that is why punishment and sadism, though they can be intertwined,[53] differ radically. Punishment is always predicated on the prior behavior of the target, and that prior behavior is always wrongful (or considered to be wrongful). Likewise, punishment has numerous purposes, but the infliction of sadistic delight is not among them. Thus, the assertion that punishment, if it does not achieve the goal of compliance but is administered nonetheless, is sadistic defies not only common usage but also, it could be argued, defies common sense.[54]

---

[52] Baldwin, *Economic Statecraft*, p. 182, footnote 105.

[53] For example, one could envisage a situation in which the individual meting out punishment to an offender derived pleasure from the offender's pain. But pain that is inflicted in order to derive sadistic delight is not, by definition, punishment.

In short, much of the literature on sanctions, even if it ignores the careful distinctions Hobbes was trying to make, treats the issue of punishment from a fundamentally Hobbesian perspective: because "punishment" cannot by definition exist in international politics, the term does not require much care in either definition or usage. Does it follow, however, that because our understanding of "punishment" tends to be so tied to legitimate authority, foreign policymakers cannot be motivated by *punitive intent* in their relations with other states? Certainly, for all of the essential ugliness of the milieu in which they regularly operate, foreign policymakers do not seem to be immune to having their moral sensibilities shocked by acts they regard as "wrongful." If we relax the Hobbesian requirement of a sovereign authority to distinguish "hostility" from "punishment" and if we adopt Hugo Grotius's succinct definition of punishment as "an evil suffered that is inflicted for an evil committed,"[55] it could be suggested that when policymakers are confronted with wrongdoing, primary among their considerations in formulating a response will be the desire to punish—in other words, to respond to "evil" with "evil." To see the instrumental utility of international sanctions as a punishment of wrongdoing, we must first explore the rational and the not-so-rational purposes of punishment.

Punishment and its purposes

Punishment of wrongdoing, I suggested above, is always a purposive and instrumental act. What, then, are the instrumental purposes of punishment? In a domestic context, legal harms are inflicted upon offenders for one or more of three broad purposes: for prevention, for compulsion, and for retribution. First, a harm may be meted out by public authorities in order to deter or prevent[56] future wrongful behavior by the individual being punished, or in order to deter wrongful behavior by other individuals in the community, or both. Second, and less frequently, punishment may be inflicted in order to compel an offender to cease wrongful behavior. The punishments inflicted on those refusing to comply with a lawful order (for example, the order from a judge to answer a query or from a legislature to return to work) are "compellent"[57] in nature: the harm is inflicted until the offender obeys. Third, punishment may be inflicted for retribution[58]—the infliction of pain on an offender in return for an evil inflicted on the community. In other words, the harm is seen as the appropriate response of the community to someone who had the choice to act otherwise but chose to act wrongfully.

---

[54] The "punishment-is-sadism" position tends to unravel when applied to ordinary cases of punishment for noncompliance. For example, according to this position, a librarian who continued to fine me after I refused to return an overdue book or a judge who continued to imprison me for repeatedly refusing to comply with a judicial order would have to be considered a sadist.

[55] Hugo Grotius, *The Law of War and Peace (De jure belli ac pacis)*, trans. Frances W. Kelsey (1925; reprint, Indianapolis, Ind.: Bobbs-Merill), p. 462. Kelsey translates "*malum passionis quod infligitur ob malum actionis*" as "an evil of suffering which is inflicted because of an evil of action."

[56] There is, of course, a difference. Punishments can *prevent* an offender from repeating the offense by making the consequences of the transgression costly enough that the offender will seek to avoid experiencing these costs in the future; by rehabilitating the offender through penitent expiation, catharsis, or correction; or by making it physically impossible for the offender to repeat the crime by amputation, lobotomy, or removal of the offender from the

It will be immediately evident that the first two purposes of punishment involve an essentially rational means-ends calculus. The harm is inflicted as a means to achieve an identifiable end that is always future-oriented. Harm is inflicted on an offender in order that something *will* occur: (1) the offender will comply; (2) the offender will be deterred or prevented from offending again; or (3) others will be deterred. The retributive purpose of punishment, by contrast, involves a minimally rational calculus of the purposes of punishment. The retributivist's reasoning is backward-looking and predicated only on what *has* occurred: (1) the offender should be punished because he broke the law (following J. D. Mabbott[59]); (2) the offender should be punished because he did wrong, and a wrong must therefore be done to him (the more traditional response that dates back to the *lex talionis* of ancient civilizations but continues to inform how most political communities view punishment[60]); or (3) the offender must be punished because we cannot allow such a wicked act to go unpunished; if we did, we would be endorsing evil (another variant of retributivism[61]). There is nothing in the retributive punishment that is directed toward the future actions of either the offender or others.[62] Indeed, the reasoning, such as it is, circulates entirely around the offense. For what reason is the offender being punished? Because he offended. What purpose will be served by punishment? The offender will be harmed and will thus receive a just desert for the offense. Moreover, by harming the offender,

---

community through incarceration, banishment, or execution. But only the first of these preventive punishments is, properly speaking, a "deterrent punishment," for it makes little sense to argue that an offender who does not offend again because he has been rehabilitated or because it is physically impossible for him to do so has been "deterred" from offending again.

[57] This companion term for deterrence was coined by Thomas Schelling in *The Strategy of Conflict* (New York: Oxford University Press, 1960), p. 195.

[58] This may, but need not, include the purpose of restitution: imposing on the offender responsibility for restoring the *status quo ante malum*.

[59] Mabbott argues that punishment is the necessary and obligatory corollary of law breaking, not wrongdoing: one punishes a person not because he is evil, but because he has broken a law. See J. D. Mabbott, "Punishment as a Corollary of Rule-Breaking," in Rudolph J. Gerber and Patrick D. McAnany, eds., *Contemporary Punishment: Views, Explanations and Justifications* (Notre Dame, Ind.: University of Notre Dame Press, 1972), pp. 41–48.

[60] For example, the durability of the retributive justification for punishment can most clearly be seen in modern public policy debates about the appropriateness of capital punishment.

[61] For a discussion, see Feinberg, *Doing and Deserving*, pp. 103–13.

[62] See the discussion in R. A. Duff, *Trials and Punishments* (New York: Cambridge University Press, 1986), p. 236.

we will have avoided becoming *particeps criminis*. Although there is no clear future-oriented end in such a formulation, this does not suggest that the punishment is neither purposive nor instrumental. Retributive punishment may be motivated by inchoate reasoning[63]; it may be nonrational; it is impossible to justify normatively.[64] But punishment for retribution is no less purposive or instrumental than punishments imposed to deter or compel. The "useful purpose" of this punishment is the harm that will be done to the offender, an equivalence of evil suffered for an evil committed—an analytical construct that is fundamentally different than "petulance," "punishment for its own sake," or sadism.

It will also be evident that deterrent and compellent punishments, because they are future-oriented, are susceptible to "success" or "failure." The subsequent behavior of the offender (or others) will determine whether inflicting a particular harm on an offender has prevented, deterred, or compelled, and thus the "effectiveness" or "ineffectiveness" of deterrent and compellent punishments will become evident—in the fullness of time. In the case of retributive punishment, by contrast, there is both an automaticity and an immediacy to its "success" and "effectiveness." Because it is imposed without regard to future behavior, such punishment is immediately "effective"—as soon as, and to the extent that, the offender is harmed by the punishment. In short, retributive punishment, by its very nature, always "works."

Punishment and sanctions: the case of Afghanistan

What tends to be missing from orthodox discussions of the useful purposes of international sanctions is a recognition that these measures, as an instrument of punishment for wrongdoing, may be motivated by the same nonrational and inchoate calculus that underlies punishment in a domestic context. It was suggested above that when foreign policymakers are confronted by what they regard as wrongdoing or evil behavior by other governments, they are likely to be moved by the same desire to punish the perpetrators that moves us to punish individuals who commit wrongful or evil acts in civil society. And just as that punitive urge may have a number of rationales in a domestic context, so too might it be argued that various rationales are possible in punishment at the international level. The reaction of the United States to the Soviet invasion of Afghanistan provides an instructive case study of the punitive motive at work.

---

[63] As Feinberg notes in *Doing and Deserving*, pp. 103–4, the notion that a society which fails to punish wicked arts endorses them "does seem to reflect, however dimly, something embedded in common sense."

[64] See, for example, Honderich, *Punishment*, chap. 2, pp. 22–51; and the selections in Gerber and McAnany, *Contemporary Punishment*, chap. 2.

Because the Soviet Union did not withdraw its troops from Afghanistan immediately after the imposition of sanctions, this episode has frequently been held up as a good example of the "failure" of sanctions,[65] a conclusion that is inevitable if one begins with such a one-dimensional view of the effectiveness of sanctions. On the other hand, there are more sensible approaches to assessing both the purposes and the effectiveness of these measures. The sanctions could be interpreted as having been motivated in part by a desire to signal resolve and leadership to the domestic public—not an unimportant consideration in an election year.[66] Alternatively, the American response could be analyzed as a Baldwinian "influence attempt" or a Knorrian "power play," with the economic (and other) sanctions having been motivated by numerous external and domestic political purposes, including some purposes (such as "signaling" to the Soviet Union, to American allies, and to other states) having little to do with the punishment of wrongdoing.[67]

However, given the widespread perception, particularly in the West, that this invasion was a morally wrong act, can we be certain that the sanctions were not primarily driven by a desire to punish the perpetrator of that act, as, for example, Robert Paarlberg and Peggy Falkenheim, suggest?[68] To suggest the primacy of other nonpunitive alternatives, as Baldwin does,[69] is not necessarily to deny the punitive hypothesis.[70] If a desire to punish was indeed one of the motivations of foreign policymakers in Washington in this case, it could be argued that one or more of the purposes of punishment underlay the hurts administered against the Soviet Union. Three possibilities can be suggested: punishment to compel, punishment to deter, and punishment for retribution.

Punishment to compel

As Hufbauer and Schott see it, the sanctions were imposed to compel: the hurt was designed to secure a Soviet withdrawal from Afghanistan. There is little to commend this argument. While there is some evidence to suggest that the continued application of sanctions eventually became linked in the public statements of American officials to "compliant" behavior by the Soviet Union,[71] there is little evidence that anyone in the administration entertained a serious expectation that the Soviet Union would actually bow to Western sanctions and withdraw.[72] Not only were the measures unlikely to succeed in harming the Soviet Union sufficiently for that purpose,[73] but more important, the manner in which they were imposed virtually guaranteed that the Soviet leadership would never suffer the humiliating loss of face that bowing to the publicly announced American ultimatum would have entailed.[74]

---

[65] Hufbauer and Schott, for example, assign a score of "outright failure" on their "success" scale in *Economic Sanctions Reconsidered*, p. 664. For a survey of the generally negative assessments of the "effectiveness" of these sanctions, see Baldwin, *Economic Statecraft*, pp. 267–68.

[66] See Peggy L. Falkenheim, "Post-Afghanistan Sanctions," in Leyton-Brown, *Utility of International Economic Sanctions*, pp. 107 and 127.

[67] See Baldwin's cogent discussion of the "targets" of the Afghanistan embargo in *Economic Statecraft*, pp. 261–68.

[68] See Robert L. Paarlberg, "Lessons of the Grain Embargo," *Foreign Affairs* 59 (Fall 1980), p. 155: "The purpose . . . was simply to punish the Soviet Union for its invasion of Afghanistan." See also Falkenheim, "Post-Afghanistan Sanctions," p. 107.

[69] See *Economic Statecraft*, pp. 264–66, for Baldwin's view of the punitive motive in the Afghanistan case.

[70] Indeed, because the task of foreign policy analysis is to reconstruct decisions in terms that sound both purposive and rational, there is always the possibility that the "rational reconstruction" of decisions by a particularly clever analyst may result in an *ex post facto* rationalization that reads more into motivations than may have been there.

Punishment to deter

The harms imposed by the American sanctions could be interpreted as a deterrent punishment directed primarily against the Soviet Union to deter Soviet leaders from future wrongdoing of a comparable kind. Or, alternatively, the sanctions may have been a deterrent punishment imposed on the Soviet Union but aimed at all states in the international system for the purpose of demonstrating the high costs of invasion and thereby deterring them from violating a cardinal principle of the contemporary international order.

There is little evidence to suggest that the Carter administration sought to injure the Soviet Union in order to impress other states of the need to uphold the principle of the nonuse of force. Not only is there nothing in the public statements of senior officials during this period to suggest that this was a factor, but the argument itself makes little intuitive sense. The use of force by either the United States or the Soviet Union is, because of the system-determining status and power of these states, always a *cas unique;* to suggest that sanctions were imposed against the Soviet Union primarily to deter smaller states from using force by demonstrating the costs attached to such statecraft stretches credulity. Moreover, the inconsistent response of President Carter to the use of force by other states in 1979 (Vietnam's invasion of Kampuchea, China's invasion of Vietnam, and Tanzania's invasion of Uganda) prompts some skepticism about the administration's attachment to the norm of the nonuse of force.

---

[71] For example, on 11 January 1980, Secretary of State Cyrus Vance claimed that one of the purposes of the sanctions was "to make it very clear to the Soviet Union that they will continue to pay a heavy price as long as their troops remain in Afghanistan." *The New York Times*, 12 January 1980, p. 4.

[72] See, for example, Vance's pessimistic assessment of the likelihood that the Soviet Union would withdraw by the middle of February, the date set by President Carter for the imposition of a boycott of the Olympics. *The New York Times*, 21 January 1980, p. A4.

[73] For one assessment of the impact of these sanctions, see Falkenheim, "Post-Afghanistan Sanctions," pp. 111–27. Compare Hufbauer and Schott, *Economic Sanctions Reconsidered*, pp. 655–65. It might be noted that eventually the United States embraced measures that did make it more costly for the Soviet Union to continue occupying Afghanistan. But the American military assistance funneled to the Afghan rebels after 1984 was not, properly speaking, a punishment.

[74] This argument has been made regarding the Polish crisis of 1981–82. See Paul Marantz, "Economic Sanctions in the Polish Crisis," in J. L. Black and J. W. Strong, eds., *Sisyphus and Poland: Reflections on Martial Law* (Winnipeg: Ronald P. Frye, 1986), p. 122. Indeed, one might conclude that trying to compel the Soviet Union by employing sanctions will always be a mug's game: with the standard having been set by Nikita Khrushchev's ignominious descent into nonpersonhood after he bowed to the ultimatum issued by John F. Kennedy during the Cuban missile crisis, it is highly unlikely that any conceivable package of economic sanctions would move another Soviet leader to suffer comparable humiliation and risk a similar fate.

On the other hand, there is considerable rhetorical evidence to suggest that these measures were designed as punishments to deter the Soviet Union itself from further uses of force. From the outset, the public statements of senior policymakers frequently stressed that the invasion posed a threat to other states in the area. For example, Donald McHenry, the U.S. representative at the United Nations, told the Security Council on 6 January 1980 that a strong response was needed because "no state will be safe against a larger and more powerful neighbor if the international community appears to condone the Soviet Union's armed intervention."[75] Likewise, on 11 January, Cyrus Vance, the Secretary of State, stated that one of the purposes of sanctions was "to make sure that they [the Soviets] understand that aggression will be faced up to whenever it occurs."[76]

This theme was particularly evident in President Carter's statements, beginning with his first statements on the issue on 27 and 28 December 1979.[77] In a major television address on 4 January 1980, he noted explicitly that the invasion posed a threat to Iran and Pakistan and was a "steppingstone" to "possible control" of the Persian Gulf. The deterrent theme, with its implicit reference to the "Munich syndrome," was sounded most vigorously in the conclusion of that address: "History teaches perhaps very few lessons. But surely one such lesson learned by the world at great cost is that aggression unopposed is a contagious disease."[78] Likewise, the State of the Union address on 23 January made clear the link between the injury of sanctions and deterring further Soviet military expansion.[79] The punishment-for-deterrence theme, not coincidentally, is repeated several times in Carter's memoirs. For example, he noted in his diary on 3 January that "unless the Soviets recognize that it has been counter-productive for them, we will face additional serious problems with invasion or subversion in the future."[80]

However, as noted above, the hurts imposed by these measures were very limited in nature, casting some doubt on the conclusion that the U.S. government was in fact hoping that the primary result of these measures would be the deterrence of similar Soviet wrongdoing elsewhere. Indeed, there is some evidence that some officials openly doubted whether the sanctions would have any deterrent effect.[81] More important, however, as David Baldwin has noted, by mid-January sanctions had been abandoned as the primary means of deterrence and had been replaced by the threat delivered in the Carter Doctrine that a comparable forceful Soviet expansion in the Gulf would be met with the use of force. Such a threat, because it raised the stakes hugely, was a far more effective means of curbing such adventuristic predilections as the Soviet leadership may have been entertaining.[82]

---

[75] *The New York Times*, 7 January 1980, p. A1.

[76] Ibid., 12 January 1980, p. 4.

[77] Ibid., 29 December 1979, p. A5; and 30 December 1979, pp. 1 and 10.

[78] Ibid., 5 January 1980, p. 6.

[79] Ibid., 24 January 1980, p. A12.

[80] Jimmy Carter, *Keeping Faith: Memoirs of a President* (New York: Bantam Books, 1982), p. 473. See also p. 471: "The brutality of the act was bad enough, but the threat of this Soviet invasion to the rest of the region was very clear."

Punishment for retribution

There is, however, a third possibility. It is undeniable that the Soviet intervention was greeted by anger in the United States[83]—and elsewhere. Given this, it can be argued that the punishment of the Soviet Union was also motivated by a retributive, and hence minimally rational, calculus. Such a retributive formulation, if it were to be made explicit, would run roughly as follows: "The Soviets have chosen to commit an evil act, and therefore some evil (such as the hurts administered by economic and other sanctions and also the reprobation that will accompany these concrete measures) will be the just desert of such an offense against the international community. Such hurts as we are willing to mete out will probably not be sufficiently harmful to deter either the Soviet Union or others from using force in defense of their interests in the future (for that, we will have to draw a line and threaten to use force if they cross it). They definitely will not compel the Soviet leaders to withdraw their forces from Afghanistan. But the act was wrongful and offensive to the international community, and because of that (and for no other reason) it cannot go unpunished. Sanctions, by harming the Soviet Union in a concrete and material manner, will achieve this end: an evil of some equivalence will have been imposed." Such a formulation, it might be noted, suggests a rationale that is both purposive and instrumental—and, more important, is quite "effective."

The public justifications offered by the decision-makers themselves for imposing sanctions suggest such a conclusion. For even if read in the wider context of the speeches or interviews in which they occur, the excerpts below are cast in the idiom of retributive, not compellent or deterre punishment:

---

[81] For a report on the debate within the administration on the effects of sanctions, see *The New York Times*, 6 January 1980, p. 16.

[82] See Baldwin, *Economic Statecraft*, p. 264–65.

[83] For example, President Carter let it be known that he was "personally angered" not only by the intervention but also by Leonid Brezhnev's response. See *The New York Times*, 4 January 1980, p. 6. For a broader discussion of Carter's images of the Soviet Union, see Jerel A. Rosati, *The Carter Administration's Quest for Global Community: Beliefs and Their Impact on Behavior* (Columbia: University of South Carolina Press, 1987), especially chap. 3.

> *President Carter, 31 December 1979:* [It is] imperative . . . that the leaders of the world make it clear to the Soviets that they cannot hav taken this action to violate world peace . . . without paying severe political consequences.[84]

> *President Carter, 4 January 1980:* The world simply cannot stand by and permit the Soviet Union to commit this act with impunity.[85]

> *Deputy Secretary of State Warren Christopher, 6 January 1980:* We intend to carry them [the sanctions] out for long enough so that the Soviet Union incurs some cost for this really quite unprecedented action. . . . We don't intend to set any time for withdrawing the sanctions. I think one of the lessons coming out of the Czechoslovak crisis is that the response needs to be determined and of considerable duration.[86]

> *Secretary of State Cyrus Vance, 15 January 1980:* I think they took a conscious, calculated decision in invading Afghanistan and they are going to have to pay the price for it.[87]

> *President Carter, 20 January 1980:* We must make it clear to the Sovie Union that it cannot trample upon an independent nation and at the same time do business as usual with the rest of the world. We must make clear that it [the Soviet Union] will pay a heavy economic and political cost for such aggressions.[88]

> *President Carter, 23 January 1980:* But verbal condemnation is not enough. The Soviet Union must pay a concrete price for their aggression. . . . That is why the United States has imposed stiff economic penalties on the Soviet Union.[89]

> *President Carter, memoirs, regarding "punitive action" against the Soviet Union for having "acted outrageously":* I was determined to lead the rest of the world in making it as costly as possible. . . . I was determined to make them pay for their unwarranted aggression.[90]

Likewise, *The New York Times* noted in its analysis of the differences within the administration about the likely effects of sanctions: "Despite these differences there is general agreement that the situation requires a credible American response."[91] In short, it could be inferred from the nature and tone of these quotations that even if the Soviet Union had bowed to American pressure or even if the leadership in Moscow actually had intended to press on militarily to the Gulf (or used force in Poland) and were deterred from doing so, there would nonetheless have been a desire to inflict harm on the Soviet Union for having acted wrongfully in the first place.[92]

---

[84] *The New York Times*, 1 January 1980, p. 4.

[85] Ibid., 5 January 1980, p. 6.

[86] Ibid., 7 January 1980, p. A6.

[87] Ibid., 16 January 1980, p. A14.

[88] Letter to Robert J. Kane, president of the U.S. Olympic Committee, quoted in *The New York Times*, 21 January 1980, p. A4.

[89] *The New York Times*, 24 January 1980, p. A12.

[90] Carter, *Keeping Faith*, pp. 471, 472, and 476.

It is true that to argue that one of the purposes of imposing sanctions against the Soviet Union was to exact a hurt for the invasion of Afghanistan may sound nonrational, particularly to the analyst who will demand to know what purpose (other than sadistic delight) will be achieved by hurting. That, however, is precisely the point: punishment, as a purposive human activity that is intended to hurt people, is, for all its purposiveness, not an activity that is amenable to high standards of means-end rationality.

Conclusion

The argument advanced here must be suggestive and heuristic, if only because of the difficulties of demonstrating conclusively that a punitive calculus, particularly in its retributive form, informs foreign policy decisions on sanctions.[93] What I have sought to show, however, is that while sanctions may have numerous other "purposes," the desire to punish will always be an integral factor in their imposition. If these measures are seen as the harms imposed for an act of wrongdoing, as I have argued they should be, then the punitive purpose of sanctions should not be overlooked or dismissed. And if we admit that sanctions are a form of "international punishment" for wrongdoing, then we cannot avoid dealing with the various reasons why human beings purposively harm each other in this way. I have argued that a model of means-end rationality has limited utility for understanding why

[91] *The New York Times*, 6 January 1980, p. 16. It might be noted that Henry Kissinger sounded a comparable theme, according to James Reston. See "Kissinger on Afghan Crisis," *The New York Times*, 4 January 1980, p. A23: although Kissinger was opposed to "grand gestures," he did agree that the Soviet Union "must be made to pay a price for its latest outrage."

[92] The comparison with the punishments invoked after the shootdown of Korean Air Lines flight 007 is instructive. Even though it was highly unlikely that such an incident would recur, there was nonetheless a desire to inflict a penalty on the Soviet Union for its action. See Hufbauer and Schott, *Economic Sanctions Reconsidered*, pp. 738–43; Hersh, *"The Target Is Destroyed,"* pp. 185–86; and Alexander Dallin, *Black Box: KAL 007 and the Superpowers* (Berkeley: University of California Press, 1985), pp. 93–94.

[93] Chief among the empirical difficulties is knowing what passes through the minds of decision-makers when they are confronted with actions of others which deeply offend their moral sensibilities. We can make some educated assumptions about the empathetic responses of a leader when informed of some act of wrongdoing; but in public statements and memoirs, the visceral and more nonrational responses will inevitably be filtered, and often mitigated, by the necessity of appearing to be acting in accordance with the precepts of means-end rationality.

punishment is imposed in a domestic context. It is argued here that it is ju as difficult to understand international harms imposed for wrongdoing if we view them in purely rational terms.

The argument that a desire to punish wrongdoing can motivate foreig policymakers brings us back to the question posed by so many students o international sanctions: Why do policymakers persist in using what is supposed to be an ineffective tool of statecraft? How we answer this questio as Baldwin rightly notes, depends on how we define effectiveness. Baldwin's approach leads us to conclude that sanctions are in fact a most “effective and rational instrument of diplomacy. And if we look at sanctions from the perspective of punishment, we come readily to the same conclusion. However, while Baldwin argues that effectiveness must be deduced from using a model of means-end rationality, I argue here that effectiveness can also be deduced from the nonrational, but nonetheless instrumental, purposes of punishment. In the case of sanctions against Afghanistan, the retributive purpose was clearly in evidence. And if economic sanctions are indeed imposed for retributive reasons, rather than for deterrent or compellent reasons, their “effectiveness” is ensured, for their effectiveness lies in their capacity to impose some harm on the target. To the extent that sanctions impose an injury on the wrongdoer (and no analyst of sanctions claims that these instruments of economic statecraft do not hurt the target in *some* way), international sanctions cannot but “work.” This may explain why sanctions have been and are likely to remain a durable and attractive policy option for foreign policymakers who are confronted with acts they regard as morally repugnant, the conventional wisdom about the ineffectiveness of this instrument of statecraft notwithstanding.

34

WHARTON REPROGRAPHICS

WORLD ECONOMY 15(2), 1992

LIPPINCOTT LIBRARY
DOCUMENT DELIVERY SERVICE

'Aggressive Unilateralism' and Section 301: Market Opening or Market Closing?

Thomas O. Bayard and Kimberly A. Elliott

1. INTRODUCTION

FACED with a surging trade deficit, the reluctance of major trading partners to initiate a new round of multilateral trade negotiations, intense pressure from some sectors for import protection, and from Congress for action, the Reagan administration, in September 1985, went on the offensive with its international economic policy. Economically, the most important policy shift was adoption of an explicit dollar devaluation policy, coordinated with our major trading partners in the Group of Seven. Recognising, however, that there is a lag before exchange rate changes are reflected in trade flows, the administration simultaneously adopted an aggressive export promotion strategy in an attempt to divert attention from the import to the export side of the trade balance.

In addition to any direct trade balance benefit that might accrue, administration officials hoped this strategy would spur more active support of a liberal trade policy from export interests, at least partially offsetting the pleas for protection from import-competing interests.[1] Driven also by an increasing perception in some quarters that US trade problems were due in no small part to unfair foreign trade practices, the new strategy relied heavily on the use of threats of trade retaliation to reduce those alleged foreign barriers. This policy strand has been

---

Copyright © 1992, Institute for International Economics, Washington, DC. All rights reserved.

THOMAS O. BAYARD is Deputy Director & Research Fellow at the Institute for International Economics, Washington DC.
KIMBERLY A. ELLIOTT is a Research Fellow at the Institute for International Economics, Washington DC.

They would like to thank Ruth Junkin and Charles Iceland for their tireless research assistance, David Richardson and Douglas Nelson for their detailed and insightful comments on the paper, and Marcus Noland, William R. Cline, and Philip Verleger for suggestions on improving the probit analysis.

[1] For a discussion of trade policy pressures in the 1980s and the congressional and administration responses to them, see Destler (1992).

dubbed 'aggressive unilateralism' by Jagdish Bhagwati, who is its most prominent critic.[2] As Bhagwati defines it, the policy tends to be unilateral in two senses: (1) the United States often unilaterally determines what is unfair in others' trade practices; and (2) foreigners are expected to undertake unilateral liberalisation.

This paper examines one of the leading policy instruments of aggressive unilateralism: section 301 of the Trade Act of 1974 (as amended). Beginning as early as 1979, Congress amended section 301 to limit the President's discretion and encourage more aggressive prosecution of cases. In the 1984 Trade and Tariff Act, Congress explicitly authorised the US Trade Representative (USTR) to 'self-initiate' cases (without waiting for a petition from domestic industry). Despite the flurry of self-initiated cases in 1985-86, Congress continued its efforts to force the administration's hand on trade policy. Those efforts culminated in the 'super' and 'special' 301 provisions in the 1988 Trade Act, which required the president to specify trade liberalisation priorities, including priority foreign countries and practices that would be the subject of 301 investigations (Table 1 summarises the statutory evolution of section 301).

TABLE 1
The Evolution of Section 301

Law	*Presidential Authority*	*Change from Previous Law*
Trade Expansion Act of 1962 (Section 252)	Broad discretion to retaliate against 'unjustifiable' foreign trade barriers with greater authority to retaliate against agricultural barriers; limited authority to retaliate against 'unreasonable' barriers.	Supplemented, but did not replace Section 350 of the Tariff Act of 1930 (enacted under the Trade Agreements Act of 1934).
Trade Act of 1974 (Sections 301-302)	Expanded discretionary authority to retaliate against both unjustifiable and unreasonable foreign barriers, with no distinction made between agricultural and non-agricultural goods.	Replaced section 252; extended coverage to services 'associated with international trade;' authorised action against foreign export subsidies; required USTR to submit reports to Congress every six months
Trade Agreements Act of 1979 (Sections 301-306)	Unchanged.	Clarified applicability of statute to services 'whether or not associated with specific products;' established more detailed procedures for investigations, including time deadlines for action; required consultations with foreign trade partners and use of available dispute settlement procedures.

---

[2] For the best and most comprehensive assessment to date see Bhagwati and Patrick (1990).

Law	*Presidential Authority*	*Change from Previous Law*
Trade and Tariff Act of 1984 (Section 301-307)	Discretionary authority unchanged; but USTR now permitted to self-initiate investigations and recommend action to the president.	Explicitly authorised retaliation in the services sector; explicitly included for the first time coverage of intellectual property and direct foreign investment; required submission to Congress of National Trade Estimates report.
Omnibus Trade and Competitiveness Act of 1988 (Sections 301-310)	Authority to retaliate shifted from president to USTR, subject to the specific presidential direction, if any; retaliation against unjustifiable practices made mandatory, but with many loopholes permitting considerable discretion.	Established 'super 301', requiring USTR in 1989 and 1990 to identify trade priorities, including designating 'priority countries and practices' to be investigated under section 301; established 'special' 301 to promote more aggressive assertion of intellectual property rights; established new deadlines for action in cases involving GATT dispute settlement or intellectual property.

Advocates of section 301 argue that the GATT is weak and increasingly irrelevant, while aggressive use of section 301 has been a successful tool of US trade policy. Its critics warn that aggressive unilateralism is more likely to result in closing US markets than in opening markets abroad; they argue further that it undermines the multilateral trading system and sets unfortunate precedents that eventually could be used against the United States by the European Community (EC) and Japan.[3] This paper concludes that the critics of section 301 exaggerate its evils, but that its fans exaggerate its merits.

This paper attempts to shed light on the debate by systematically analysing all completed section 301 cases investigated since 1975 for which the outcomes are known. We find that section 301 has led to at least partial market opening in just over half the cases overall, and in two-thirds of the cases ending after announcement of the President's trade policy action plan in September 1985. The imposition of US retaliation under section 301 has been rare (only 8 of 89 investigations as of July 1992) and the incidence of foreign counter-retaliation even rarer (only one case). It should be noted, however, that the trade impact of section 301 cases typically is very small. In a quarter of the cases, total US exports (in a few cases imports) of the product in question to (from) the targeted country had a value of less than $10 million. In three-quarters, the value was less than $200 million.[4]

---

[3] On the precedential implications see Bayard (1990).

[4] While gross export figures generally will overstate the incremental trade gain that might be expected in these cases, this measure likely understates the benefit derived in some cases where the precedential or deterrent effect may be much larger than the immediate trade effect.

This paper however, examines only one part of the debate: whether the US Trade Representative achieved its negotiating objectives, narrowly defined as improved market access for US goods and services. It does not address the question of the impact of aggressive unilateralism on the functioning of the liberal multilateral trading system.[5] Nor does it address the utility of aggressive unilateralism as a market-opening tool relative to other tools; for example, multilateral negotiations. Those issues are set aside not because they are unimportant, but simply because of space constraints. Those issues, which are essential to a complete analysis of aggressive unilateralism, will be addressed in the larger research project of which this article is just one part (see Bayard and Elliott, forthcoming).

2. EXPERIENCE WITH SECTION 301 SINCE 1975

Just ten countries (counting the EC as one country) account for 75 of the 82 cases investigated from section 301's creation in 1975 through the intellectual property case against Taiwan in April 1992.[6] Half of those investigations targeted the three largest US trading partners: the EC, Canada and Japan. Interestingly, the EC rather than Japan is by far the most frequent target and agricultural disputes have dominated the US trade agenda with the Community. Just under a quarter of *all* 301 cases have involved agricultural disputes with the EC, while nearly half overall have involved disputes over agriculture. Just over a third of the cases sought access for manufactured goods, while the remainder addressed the 'new' areas of services and intellectual property (see Table 2).

It is interesting that Mexico, the fourth largest US export market, has never been the subject of a 301 investigation, though it was placed on the 'priority watch' list released in May 1989 under the 'special 301' provisions of the 1988 Trade Act. (Mexico was removed from the list the following year.) Argentina on the other hand has been a surprisingly frequent target of 301 even though it is a relatively small US export market; however, three of the five cases against Argentina involved services or intellectual property issues.

---

[5] For discussion of the relationship between section 301 and the international trading system, see Bello (1990), Hudec (1990) and Ryan (1991).

[6] There were actually 89 cases investigated by the US Trade Representative's office in this period. Two — involving imports of subsidised Canadian softwood lumber (Nos. 58 and 87) — are excluded from this analysis because section 301 was used only for administrative purposes, not as a negotiating tool. Six more (Nos. 27-31 and No. 33) are counted as one case because they had the same petitioner and involved the same products and issues. Those six complaints, regarding allegedly subsidised steel imports from six European countries, were eventually shifted into a section 201 escape clause investigation at the order of the president.

TABLE 2
Breakdown of 301 cases by Country and Product*

		Type of Product			
Target	*No. of Cases*[a]	*Mfd.*[b]	*Ag.*[c]	*Service*	*IP*
All	82 (41)	28.5 (12.5)	35 (13)	10 (7)	8.5 (8.5)
EC	22.5 (9)	2.5	19 (9)	1	0
Japan	12.5 (8)	7.5 (6)	4 (1)	1 (1)	0
Korea	8 (4)	2	3 (1)	2 (2)	1 (1)
Taiwan	7 (3)	4 (1)	2 (1)	0	1 (1)
Canada	6 (2)	4 (1)	1	1 (1)	0
Argentina	5 (3)	1 (1)	1	2 (1)	1 (1)
Brazil	5 (3)	2.5 (1.5)	1	0	1.5[d] (1.5)
India	4 (3)	1 (1)	1	1 (1)	1 (1)
Thailand	3 (3)	0	1 (1)	0	2 (2)
PRC	2 (2)	1 (1)	0	0	1 (1)
USSR	1 (1)	0	0	1 (1)	0
Others[d]	6	3	2	1	0

Notes:

* Number of cases involving public, explicit threats, or retaliation, are in parentheses. Public threats are defined as self-initiation of an investigation by USTR; publication of a presidential determination that an actionable policy or practice exists, including a definite date for taking action; publication of a retaliation 'hit list'.

[a] As of June 1992. Excludes two cases involving Canadian exports of softwood lumber. Includes six steel cases against Austria, France, Italy, Sweden, the UK, and Belgium that were eventually transferred to a section 201 case, counted here as one case and included under 'other'; another case, alleging the 1976 EC-Japan steel agreement burdened US commerce, is split between the EC and Japan as targets.

[b] Includes case brought over Taiwan's customs valuation system, the super 301 case against Brazil for its import licensing system, and the super 301 case against India for restrictions on foreign investment, and the general case against China in 1991-92. The Brazilian informatics case, which involved both market access issues and intellectual property concerns, is split between the manufactured and IP categories.

[c] In addition to raw agricultural products, includes processed food and other agricultural and natural resource-based products *if* the issue arises from a subsidy to or protection for processors to offset the high cost of protected agricultural inputs; e.g., pasta, wine, tobacco products.

[d] 'Others' include Guatemala, Portugal and Spain, prior to EC accession, and Norway, named in one case each, and those countries named in the steel cases referred to in footnote a.

TABLE 3
Breakdown of 301 Cases Over Time*

Target	*Total No. of Cases*	*Type of Product*			
		Mfd.[h]	*Ag.*[a]	*Service*	*IP*
1975-79	21 (6)	5 (1)	11 (2)	5 (3)	
EC	8.5 (2)	0.5	8 (2)		
Japan	4.5 (1)	2.5 (1)	2		
Other[b]	8 (3)	2	1	5 (3)	
1980-84	21[c] (5)	10 (2)	8 (2)	3 (1)	
EC	6 (2)	1	4[d] (2)	1	
Japan	1 (1)	1 (1)			
Taiwan	3	2	1	0	
Other[e]	11 (2)	7 (1)	3	1 (1)	
1985-89	31 (23)	10[f] (8)	14 (8)	3 (3)	4 (4)
EC	6 (4)	1	5 (4)		
Japan	7 (6)	4 (4)	2 (1)	1 (1)	
Korea	5 (3)		3 (1)	1 (1)	1 (1)
Brazil	3 (3)	1.5 (1.5)			1.5 (1.5)
India	3 (2)	1 (1)	1	1 (1)	
Other[g]	7 (5)	3 (2)	3 (2)		1 (1)
1990-91	9 (7)	2 (1)	2 (1)		5 (5)
EC	2 (1)		2 (1)		
Thailand	2 (2)				2 (2)
Other	5 (4)	2 (1)			3[h] (3)

Notes:
* See notes to Table 2.
[a] Includes two against Canada, and one each against Argentina, Korea, and Taiwan.
[b] See table 5.1, note b.
[c] Includes six steel cases against several European countries that were eventually transferred to section 201 case; counted as one here and under "other".
[d] All filed in 1981.
[e] Includes two each for Argentina, Brazil, Canada, and Korea.
[f] Includes case brought over Taiwan's customs valuation system, super 301 against Brazil for its import licensing system, and super 301 against India for restrictions on foreign investment.
[g] Includes two each for Argentina and Taiwan.
[h] Special 301 designations of India, China, and Taiwan.

The prominence of agriculture on the 301 agenda is consistent across the period studied. However, Table 3 shows the increasing importance accorded to intellectual property protection in recent years. In contrast, services, the other major 'new area' being negotiated in the Uruguay Round has been the subject of investigations less often in recent years than in the late 1970s. Table 3 also documents the more aggressive trade policy adopted in the second Reagan administration. The number of 301 investigations increased by 50 per cent in the 1985-89 period relative to the previous five-year period, with intellectual property cases alone accounting for 40 per cent of the increase. Even more striking is the sharp increase in the proportion of cases involving public threats (including cases self-initiated by USTR; these cases are shown in parentheses in Table 3). Self-initiated cases occurred in two waves — one in 1985-86, when the Reagan administration was trying to influence the debate in Congress over the omnibus trade bill; and the other (with a pause in the middle) in 1989 and 1991 when the Bush administration had to implement the super 301 and special 301 provisions included in the 1988 Trade Act.

3. COMPETING AND COMPLEMENTARY HYPOTHESES ABOUT HOW SECTION 301 WORKS

Although the United States typically makes unilateral demands of its trading partners in section 301 cases, it rarely makes take-it-or-leave-it offers. The outcome is still a negotiated one, with the US negotiators trying to get as much liberalisation as they can, and foreign negotiators trying to give up as little as possible. Since the US typically is not offering reciprocal market-opening concessions in these negotiations, its leverage derives from the threat, implicit or explicit, to close its own market. Thus, the critical elements in the success or failure of a section 301 negotiation are, on the one hand, the value that the target country places on maintaining access to the US market and the credibility of the US threat to retaliate on the other.

Theories of bargaining and games suggest that threats typically 'succeed' when the economic and political costs to the target of complying with a demand is lower than the costs of defiance (see for example, Cline, 1982; McMillan, 1990; and Sykes, 1990). The costs of defiance depend on the likely size of any US retaliation and the probability that it would actually be imposed if negotiations break down. The economic costs of compliance are essentially the adjustment costs that would be borne by import-competing sectors if trade barriers are liberalised. Probably of more immediate concern to the target country government is the political pain that it anticipates could be inflicted by those and other interests in the target country that oppose liberalisation.

John McMillan has used game-theoretic models to illustrate how section 301 can 'shift the terms of agreement in the United States' favor.' He concluded that

> This shift will be larger (a) the greater the harm to the targeted country from having its access to the US market limited; (b) the smaller the targeted country's ability to harm the US in retaliation; (c) the smaller the costs within the targeted country of complying with the US demands; and (d) the greater the benefit to the United States — in the US negotiators' perception — from the demanded liberalization (McMillan, 1990, p. 213).

Obviously none of these factors is easily measured. In particular, the potential political gains and losses arising in each case would be difficult to quantify and the research required to do so for each of the 67 cases is prohibitive.[7] The data requirements for precise measurement of the potential economic gains and losses to both parties are similarly difficult given the number and nature of the products and countries involved.

One piece of data that is available for most of the cases is the actual value of US exports (imports) of the product in question to (from) the target country.[8] Although obviously flawed, this measure is expected to provide at least an order of magnitude for the level of stakes involved in each case. Unfortunately, in the absence of any adjustment for the political weight of affected constituencies in the two countries, it is not clear *a priori* whether the level of trade stakes would tend to have a linear relationship with outcomes in either a positive or negative direction. If one assumes that potential US benefits from liberalisation increase with the value of current exports, one would expect increased success to be associated with increases in the value of this variable.[9] However, if adjustment costs from liberalisation in the target country also increase with the current value of US exports, one would expect a negative correlation between this variable and success. Thus, no prediction can be made as to the expected sign on this variable; one might even expect that the coefficient would be insignificant because of the offsetting effects in the two countries.

As pointed out by McMillan, each party's bargaining strategy will also be affected by their perceived vulnerability to retaliation or counter-retaliation. In what follows, the target country's vulnerability to US retaliation is represented by the proportion of the target's total exports sold in the US market.[10] It is expected that the higher this ratio, the higher the odds of a successful outcome for US negotiators. US vulnerability to counter-retaliation is represented by a dummy variable that is equal to 1 if the target is either the European Community or Canada, and 0 otherwise. The choice of this variable is based on the assumption that these are the only targets to which the US attaches a positive probability of counter-retaliation, and which have enough clout to inflict significant pain if they do so. This assumption derives from the fact that the EC is the only US trading partner to counter-retaliate in the context of a section 301 case, while both the EC and Canada have 'retaliated' (taken 'compensatory withdrawal' of concessions granted to the United States) in the context of escape clause cases; both also regularly threaten to counter-retaliate in trade disputes with the United States.[11]

---

[7] The relative importance of domestic industry support for USTR in its negotiations, and the role of political constituencies for and against liberalisation in the target country are issues to be explored in the limited number of detailed case studies that we plan to do as part of our broader research into reciprocity and retaliation in US trade policy; see Bayard and Elliott (forthcoming). Also see Odell (1992).

[8] In some cases, where the data does not exist or where the actual value of exports is not appropriate (for example, Japanese leather and nonrubber footwear, where a quota severely restricted exports and ex-post trade cannot be observed because the barrier was never lifted), other estimates had to be used, usually either government or industry estimates of lost exports.

[9] This assumption is based on the expectation that the incremental increase in exports from liberalisation will be larger the larger is the current base value of exports.

Tactics also can enhance a negotiator's credibility and increase bargaining leverage. Schelling (1960) has stressed the importance of commitment in making a threat credible. For example, failure to carry out a public threat can have worse consequences for a negotiator's reputation, thus weakening his or her effectiveness in future negotiations, than not carrying out an implicit or private threat. In the 301 context, announcement of a presidential finding of an unfair practice, with a specific deadline for action, and release of a retaliation 'hit list' by USTR are defined as public and explicit threats.[12]

Self-initiation of section 301 investigations (relevant only since 1985) could have a similar effect by signalling that the US Trade Representative places a high priority on a case. In the debate over the omnibus trade bill in 1987 and 1988, the Reagan administration argued that making self-initiation of some cases mandatory would lessen the impact of self-initiation since such cases 'currently had clout because they were extraordinary' and signalled to target governments

---

[10] Some observers have suggested that the share of exports to the United States in the target country's gross domestic product would be a better measure, often citing India as a case where an export share may overstate its dependence on the US market since exports are relatively unimportant to its economy. But this alternative measure may also understate such dependence in some cases where exports are critical to generate foreign exchange (either to pay for needed imports or to pay off debts), whatever the share of trade in the economy. Moreover, for every India, where making GDP the denominator lowers the ratio, there is just as likely to be a Taiwan, where using GDP raises the ratio.

[11] US exports to the target as a share of total US exports might be considered as a proxy for US vulnerability to counter-retaliation. But this variable could also reflect US interest in a market, resulting in a positive correlation with success rather than the expected negative one. Visual inspection of the data does indicate a positive relationship up to a level of about 15 per cent of US exports, when it turns negative. As it happens, the only targets above that level, for the period of the sample, are the EC and Canada.

[12] As many observers have pointed out, implicit threats, including the threat to initiate a 301 investigation, may be as effective, or even more so, than public threats. Unfortunately, private and implicit threats often are unobservable and thus cannot be consistently measured or systematically modelled.

'that the administration meant business (see Bello and Holmer, 1990, pp. 83-85). Taking a case before a GATT panel might also signal USTR's commitment to a case, indicating the negotiators think they have a good case. In essence, making a threat public raises the costs to US negotiators of backing down and increases the probability that a threat will be carried out if a satisfactory bargain is not struck. Making a threat explicit (by specifying a hit list of products) makes it clear who will pay the costs of retaliation and raises the costs of defiance for the target country government if those sectors become politically mobilised.

A GATT panel finding of noncompliance with international rules may also increase the costs of defiance for the target country by adding international opprobrium to unilateral US condemnation, or by damaging the country's reputation for abiding by its commitments. Alternatively, use of the GATT dispute settlement process could lower the costs of compliance in some countries by providing domestic political cover for the target country government, allowing a change in policy to be interpreted as fulfilling one's international obligations, rather than as caving in to US pressure. To test the hypothesis that GATT rules make a difference in determining outcomes, we compare cases in which a GATT panel was convened with those in which there was no panel.

Finally, it might be expected that the likelihood of a negotiated agreement resulting in genuine market-opening would be affected by the type of trade barrier being discussed. Traditional border measures — tariffs and import and export quotas on goods — may be relatively easier to negotiate because they are more transparent, objectives are easier to define, and because such barriers are more likely to be clearly GATT illegal. Transparency may also make agreements to eliminate or reduce such barriers easier to monitor and enforce. Other types of barriers — subsidies, state trading, technical standards, etc. — may be less transparent and their impact harder to measure, perhaps leading to fuzzy agreements that are more difficult to enforce. In the analysis that follows, each case was assigned to one of four categories: border measures (tariffs and quotas); subsidies (domestic and export); administrative and services barriers (including state-owned monopolies, such as Thailand's tobacco monopoly, technical standards, and even more ill-defined barriers, such as in the Japanese semiconductor case); and lack of protection for intellectual property, put in a separate category because it is more likely than most of the others to require legislative as well as administrative action. The categories of trade barriers are listed in descending order of transparency and clarity and coverage of GATT rules. Thus, we would expect disputes over border measures to be the easiest to resolve, subsidies somewhat less so, with vaguer, administrative barriers and barriers to services trade, where there are as yet no international rules, to be the most difficult.[13]

---

[13] Although a separate category was created for intellectual property cases, that variable is not included in the statistical tests because there are only four such observations, too small a number on which to base meaningful conclusions.

4. DATABASE AND METHODOLOGY

Of 89 investigations initiated from 1975 through June 1992, 22 have been excluded for various reasons, leaving 67 cases to be analyzed.[14] The sample includes the six investigations that resulted from the spring 1989 'super 301' designations and they are treated the same as all other section 301 cases since the statutory procedures, once the designation has been made, are identical.[15] The sample does not include any of the cases initiated in 1991 or 1992 under the 'special 301' provisions of the 1988 Trade Act because they are either pending or were too recently concluded to fairly assess the outcome. The assessments of these cases are based on information from the US Trade Representative's public files, supplemented by press reports, interviews and surveys of US government officials, petitioners and their counsels, and others involved in the cases, as well as examination of trade data.

A 'successful' case is defined here as one in which US negotiating objectives — generally improved market access — were at least partially achieved. This definition of success obviously is a narrow one and is specific to our analysis of negotiating tactics and strategy. In some cases, the result might not be considered a success for US trade policy more broadly defined to include strengthening the liberal multilateral trade system. For example, the outcome in some cases was improved access for specific US firms; success in these cases does not necessarily mean the market was liberalised because market shares may only have been reallocated to benefit US firms.

In our view, and in contrast to some other observers, conclusion of an agreement is not sufficient to call an outcome a negotiating success. If a case recurs, because the agreement was too loosely written, was not implemented to US satisfaction, or was circumvented in some other way, it is classified as a failure. For example, case numbers 26 and 60, involving EC canned fruit subsidies and meatpacking standards, respectively, are classified as failures even though agreements were reached because both cases later recurred (case numbers 71 and 83). Though harder to identify than cases that reappear on the

---

[14] As of June 1992, nine cases were either pending or too recently concluded to make trustworthy assessments about the outcome (Nos. 80 and 82-89). In three other cases, the President issued a formal determination that the practices complained about were not actionable under section 301 (Nos. 10, 38, and 46), while in another (No. 21), no determination was made and the investigation resulted in a change in US, not foreign law. Cases 27-31 and 33 were ultimately shifted to a 201 escape clause investigation, while case No. 32 was handled as a countervailing duty case. Two Canadian softwood lumber cases (Nos. 58 and 87), initiated as a subsidy case and ultimately settled by bilateral agreement, are also excluded. Finally, one case (No. 39) is excluded because the petition was withdrawn before any determination was made and the outcome remains unknown.

[15] 'Super 301' (section 310 of the Trade Act of 1974, as amended) affects only the selection of cases by USTR, not their prosecution.

USTR list of cases, we also have designated as failures other cases in which agreements were reached if it appears that they did not lead to improved market access. For example, in the Thai cigarette case, the Thai government, after agreeing to dismantle its ban on cigarette imports, replaced the GATT-illegal ban with GATT-legal tariffs. Agreements to monitor, discuss, or negotiate in other fora that result in no change in the targeted policy or practice are also considered failures (for example, cases 16 and 18, involving EC wheat export subsidies and Argentine protection of its marine insurance industry, respectively).

In addition to doing simple tabulations to see which variables appeared to be correlated with success, we also calculated chi-square statistics to determine whether associations between our 'success' variable and other variables were statistically significant, and used the probit regression technique to try and establish the relative importance of the independent variables in determining successful outcomes.

The probit technique was chosen because it is more appropriate than OLS regression for equations with a dichotomous dependent variable such as success/failure. If we code a dichotomous dependent variable, Y, in the following manner: 1=success and 0=failure and then use regression analysis to estimate values of Y for given values of an independent variable, X, what we in effect are doing is estimating *probabilities* of $Y=1$ (success), given values of X. These probabilities must necessarily lie between 0 and 1. Although this is true *a priori*, there is no guarantee that the estimates of Y from OLS regression will necessarily fulfil this restriction. The probit model resolves this problem by restricting estimates of Y to the 0-1 interval. Another feature of the probit model is that the relationship between estimated probabilities of success ($Y=1$) and values of X is nonlinear; at very low and very high values of X, a small change in X will yield a very small change in the probability of success ($Y=1$); by contrast, at intermediate values of X, the same small change in X will yield a much larger change in the probability of success. Thus the model, instead of being a straight line as in OLS regression, is an S-shaped curve. The model is based on the normal probability distribution function, more popularly known as the 'bell-shaped curve'.

The chi square test is based upon a comparison of observed values for two variables and values that would be expected if the two variables were *statistically independent* (no relationship). A chi square test statistic is calculated. A critical value is then obtained by selecting a desired level for alpha (the probability of *falsely* rejecting the null hypothesis that no relationship exists between the two variables) and the appropriate degrees of freedom. If the chi square test statistic is larger than the critical value, one may reject the null hypothesis that no relationship exists between the two variables and conclude that the two variables are indeed related. The chi-square test results were useful in buttressing conclusions from the probit regressions concerning which variables appear to be significant in explaining outcomes of section 301 cases.[16]

5. GENERAL RESULTS

Based on the evidence at hand, our best judgment is that US negotiating objectives were at least partially achieved 54 per cent of the time overall (36 of 67 cases). Since September 1985, two-thirds of all cases have resulted in achievement of some or all of the US negotiating objectives (26 of 38) compared to just one-third from 1975 through the first part of 1985 (10 of 29) (Table 4).

TABLE 4
Factors Influencing the Effectiveness of Section 301
(number of cases, unless otherwise noted)

	Number of Successes (36 cases)	*Number of Failures (31 cases)*	*Successes/Total Number of Cases %*
Avg. US exports to target	$263 mil.	$509 mil.	
US exports to target:			
< $10 million	8	11	42
$10 million – $1 billion	27	16	63
> $1 billion	1	4	20
Avg. target export dependence	29%	18%	77
Target export dependence:			
> 30 per cent	17	5	77
< 30 per cent	19	26	42
Threat made public[a]	20	13	61
Self-initiated	13	4	76
Other public threats	12	10	55
Retaliation imposed	2	6	25
No known threat	16	18	47
GATT panel convened	11	9	55
Excluding the EC	8	3	73
No GATT panel[b]	25	22	53
Border measures targeted	19	6	76
Other types of trade barriers targeted	17	25	40
GATT case involving a border measure	10	3	77
GATT case involving other types of barriers	1	6	14
Cases ending:			
Before September 1985	10	19	34
After September 1985	26	12	68

Notes:
[a] Subcategories do not add to this total because of overlap among the various categories of types of threats.
[b] Includes cases to which GATT does not apply; for example, services and intellectual property.

---

[16] For those interested in the details of the statistical tests, see the technical appendix in Bayard and Elliott (forthcoming).

The general results are presented in this section; Section 6 discusses the results for the more recent period and possible factors contributing to the increased effectiveness of section 301 since announcement of President Reagan's Trade Policy Action Plan in September 1985.

An at least partially successful outcome is most likely when the targeted trade barrier is transparent and inconsistent with international rules, and when the target is relatively dependent on the US market. Public and explicit threats also appear to contribute to successful outcomes, though the effect is less robust than expected. Success appears to be somewhat less likely if the target is the EC or Canada, or if the value of US exports involved is either very small or very large (in this sample, less than $10 million or more than $1 billion); however, these results are not robust and should be treated with caution.

Table 4 suggests that increasing US export values has a negative impact on the likelihood of a successful outcome, with the average value of US exports in unsuccessful episodes being nearly twice the value of exports in successful cases. The coefficient for this variable in the regression (col. 1 of Table 5) has a negative

TABLE 5
Regression Results

Independent Variables	*Basic Equation*		*Substituting DUMMY for Stakes*		*Introducing TPAP and TBILL*	
Constant	−2.20	(−2.38) **	−2.74	(−2.75) *	−2.08	(−2.10) **
STAKES	−0.00005	(−0.11)				
DUMMY			0.60	(1.44) ***	0.09	(0.17)
ECCAN	−0.96	(−2.17) **	−0.81	(−1.80) **	−0.72	(−1.48) ***
TXDEP	3.97	(2.88) *	4.07	(2.78) *	4.52	(2.85) *
BORDER	2.28	(2.53) *	2.46	(2.75) *	3.15	(3.29) *
SUBSIDY	2.12	(2.12) **	2.23	(2.30) **	2.87	(2.89) *
ADSERV	0.80	(0.93)	0.90	(1.06)	0.98	(1.16)
TOUGH	1.11	(2.19) **	0.87	(1.68) **	0.14	(0.23)
TPAP					−1.56	(−2.45) *
TBILL					−0.22	(−0.38)
Per cent of cases correctly predicted		78		82		87

Note: *T*-statistics are in parentheses; using a one-tailed test
* Indicates significance at the one per cent level, ** at the five per cent level; and *** indicates significance at the ten per cent level.

Variable Definitions

STAKES	value of US exports to (imports from) target of affected product,
DUMMY	1 if stakes are between $10 million and $1 billion; 0 otherwise,
ECCAN	1 if EC or Canada is the target; 0 otherwise,
TXDEP	proportion of target's total exports sent to US,
BORDER	1 if border measure affecting goods (import and export quotas, and tariffs); 0 otherwise.
SUBSIDY	1 if a subsidy (including export subsidies); 0 otherwise,
ADSERV	1 if administrative or regulatory (including services barriers, technical standards, and state trading); 0 otherwise,
TOUGH	1 if an explicit threat and either a GATT panel ruling against target or case self-initiated; 0 otherwise,
TPAP	1 if case ended before September 1985; 0 otherwise,
TBILL	1 if case initiated after August 1988; 0 otherwise.

sign, but is insignificant. Closer inspection of the data, however, suggests that the relationship may not be linear. Substituting a dummy variable (called DUMMY in Table 5) for the continuous STAKES variable, with 1 representing cases in which the value of US exports is between \$10 million and \$1 billion, both improves the number of cases correctly predicted by the model and generates a statistically significant (at the ten per cent level) and positive coefficient for the modified variable.[17] The chi-square test also suggests there is a statistically significant and positive relationship between the DUMMY and OPENING variables (the dependent variable) (at the five per cent level in this test).

The explanation for this outcome could be that USTR does not prosecute cases with export values at either extreme as aggressively as other cases. At the high end, US negotiators may be more cautious because both the costs of carrying out a retaliatory threat and the potential costs of counter-retaliation are so high. Concerns about the possibility of counter-retaliation may have been compounded at the high end because in four of the five cases with US export values of \$1 billion or more, the target was the EC, which may be perceived as more likely to counter-retaliate. At the low end, US negotiators may simply feel that the expected gains are too small to be worth much trouble. Support for the hypothesis that USTR prosecutes these cases less aggressively is found in the fact that 90 per cent of public and explicit threats fall in the range between \$10 million and \$1 billion (versus 60 per cent of all cases) and nearly 3 of every 5 fall in a range from \$100 million to \$500 million (compared to 34 per cent of all cases).

The vulnerability of the target to US trade retaliation, as measured by its dependence on the US market for its exports, has stronger explanatory power. The average target export dependence in successful cases was 29 per cent versus 18 per cent in failures. When the target's export dependence was over 30 per cent, 77 per cent of cases were successfully resolved versus only 42 per cent of cases in which export dependence was under 30 per cent (Table 4). The chi-square test also indicates that export dependence above a threshold of 30 per cent is positively associated with success and the statistic is significant at the one per cent level. In the regression analysis, the coefficient for this variable (TXDEP) was robust and also significant at the one per cent level (Table 5). The coefficient for the dummy variable used to represent US concerns about counter-retaliation (ECCAN) is negative as expected and significant at the five per cent level in the equations reported in columns 1 and 2 of Table 5, but it is not robust and is not significant at even the ten per cent level in the chi-square tests.

The simple tabulation in Table 4, the regression results reported in Table 5, and the chi-square test results also strongly support the hypothesis that transparent border measures are the most likely to be liberalised following section 301 negotiations. Seventy six per cent of cases targeting border measures were successfully resolved versus only 40 per cent of other types of cases. The chi-square test results also show a strong, statistically significant correlation between BORDER and success; but the chi square for the subsidy variable is almost zero, suggesting no relationship with outcomes, while for the administrative and services barriers variable it indicates a negative relationship but is not statistically significant. The regression results for BORDER are also statistically significant (at the one per cent level) and robust, but in these tests, in contrast to the chi-square tests, SUBSIDY is positive and also statistically significant at the five per cent level or better while the coefficient for ADSERV is still insignificant but in this case has a positive sign. Thus, the statistical results suggest that the variables have the expected relationship to one another as well as to success (with border measures being the easiest to negotiate, subsidies somewhat less so, and administrative and services barriers the hardest), but in light of the chi-square results and the tabulation in Table 4, the coefficients on SUBSIDY and ADSERV have to be viewed with suspicion.[18]

---

[17] Iterative regressions confirm these threshholds as having the greatest explanatory power.

Unexpectedly, GATT does not appear to add much leverage: the success rate for cases in which GATT panels were convened (55 per cent) was not significantly different from that for cases in which either no GATT panel was appointed or GATT rules were not applicable (53 per cent). It should be noted that in *every* case where a GATT panel found a violation or evidence of nullification and impairment, changes in the offending policy were made, but in almost half, the target country replaced the illegal trade barrier with another type of barrier or disagreed with the US interpretation of what had been agreed. The success rate for cases involving GATT panels that also targeted border measures is significantly higher (77 per cent) than in cases where panels had to grapple with less transparent barriers (14 per cent). Unfortunately, there is no way of determining whether the strong performance of the BORDER variable is due primarily to the transparency of the barrier or to the clarity of GATT rules in those cases.

There is also a stronger positive correlation between GATT dispute settlement procedures and success if the EC is excluded. The EC appears to be a major culprit in abiding by the letter but not the spirit of the GATT and in exploiting any ambiguity in an agreement. Cases in which GATT panels were convened, other than those involving the EC, were successful 73 per cent of the time (vs. 55 per cent overall). Six of nine panel cases involving the EC ultimately were judged to be failures; in all but one case (wheat flour export subsidies, in which the panel was unable to determine whether the EC's subsidies violated the Subsidies Code), the EC changed the offending practice but exploited ambiguities in GATT and the bilateral agreements to continue to protect its agricultural producers and processors.

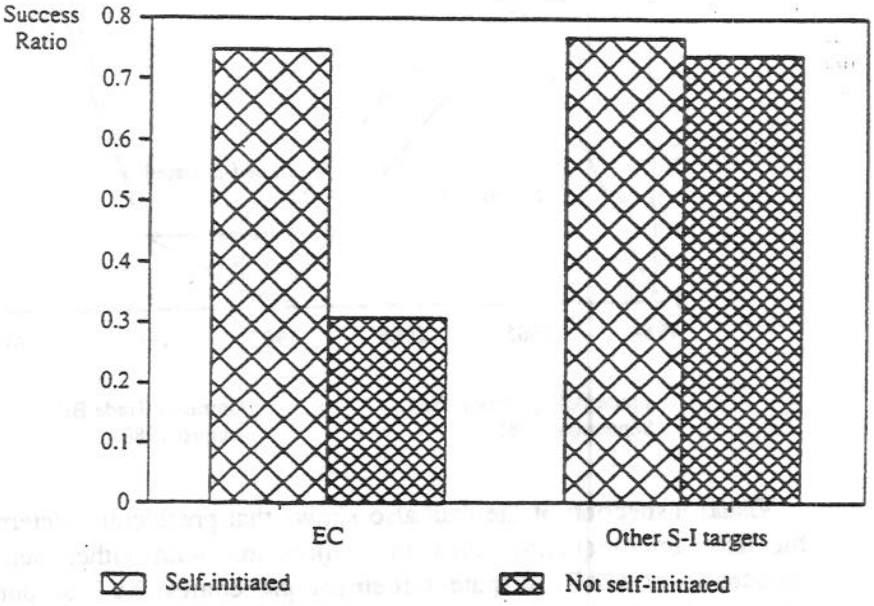
---

[18] To test for collinearity among the three types of trade variables, BORDER was dropped in one regression; the resulting coefficients on SUBSIDY and ADSERV were similar to the results from the chi-square tests. The results for BORDER were robust in any combination with the other variables. All three were left in the model because it resulted in more cases correctly predicted.

Of the four non-border measure cases that did not involve the EC and that went before GATT panels, the US agreed to terminate the proceedings in three after getting promises from the target country to change the trade practice. All three cases later recurred, indicating that USTR either did not get a good agreement or did not monitor it closely enough. Because of the ambiguity in most GATT panel reports and the complexity in interpreting the results, no attempt was made to create a variable representing GATT. As noted, the BORDER variable may also encompass clarity of GATT rules.

The statistical results provide modest support for the hypothesis that issuing public and explicit threats contribute to successful outcomes under certain circumstances. Table 4 shows that cases involving public threats are successful 61 per cent of the time, versus 47 per cent for other cases. The chi-square tests, however, suggest that the difference is not statistically significant. The variable representing self-initiated cases in the regressions was insignificant, as was the chi square, even though those cases had a 76 per cent success rate.[19] Further inspection of the data offers a possible explanation for the insignificance of the self-initiation variable, however. Chart 1a shows that the success rate for targets

CHART 1a
Self-initiated Cases: Success by Target Country



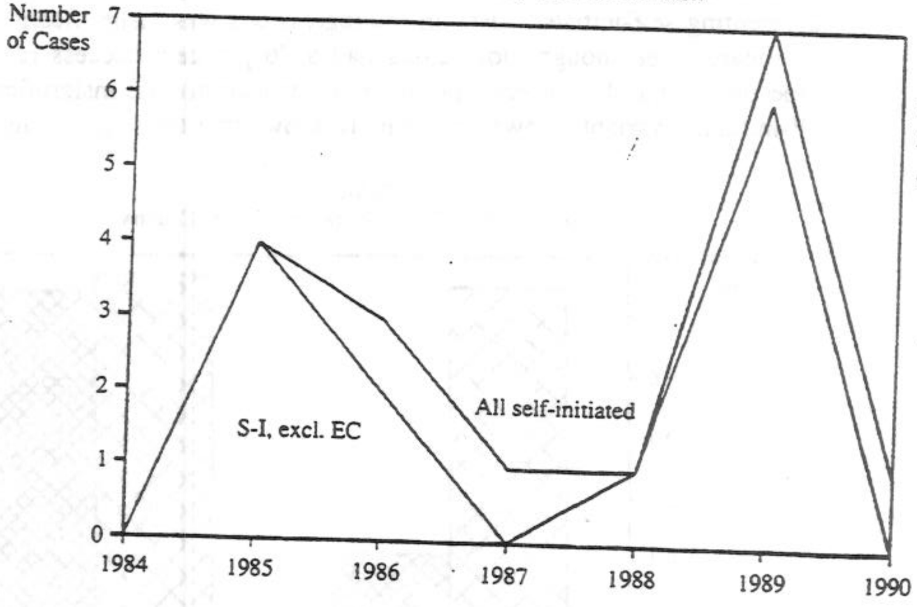
---

[19] Since self-initiation was not relevant until 1985, the sample had to be split in order to test that variable. Those results are not reported.

other than the EC that have been the subject of self-initiated investigations i identical whether or not the case was self-initiated.

Strikingly, the success rate for cases targeting the EC increases from 31 pe cent to 75 per cent if a case is self-initiated. Chart 1b shows that the use of self initiation by USTR has been highly cyclical, with a spike in 1985-86 when i was trying to head off a protectionist trade bill, and another in 1989 when it ha to implement the provisions of the Trade Act signed into law in August 1988 Although we were unable to test for the effects of domestic politics, on interpretation of the two charts is that, with the exception of the EC, USTR use self-initiation of a 301 investigation primarily for domestic political purposes rather than to send a signal to the foreign target.

CHART 1b
Self-initiated Cases by Year of Initiation



Trade Policy Action Plan,
September 1985

Omnibus Trade Bill,
August 1988

Visual inspection of the data also shows that presidential determinations and hit lists are frequently used in conjunction with either self-initiation or invocation of GATT dispute settlement procedures. So, for purposes of the regression analysis, a dummy variable (TOUGH) was created that is equal to 1 if a case is either self-initiated *or* involves a GATT panel (no self-initiated case has yet been taken before a GATT panel) *and* involves another public and explicit threat (presidential determination or hit list). The coefficient for this variable is positive and significant at the five per cent level (columns 1 and 2 of Table 5); the chi square is also significant at the five per cent level.

USTR is more aggressive in using public threats, including self-initiation, when the target is relatively vulnerable to retaliation and when the stakes are not so high as to threaten a costly trade war. Threats also appear to be used less often if a case is covered by GATT rules: only 22 per cent of cases in which GATT rules might have applied (whether or not a panel was appointed) involved public and explicit threats vs. 68 per cent of other cases. USTR is also more aggressive in section 301 cases targeting Japan and much less aggressive with the EC. Every case targeting Japan involved either a public and explicit threat or a GATT panel; slightly less than half the cases against the EC either involved threats (40 per cent) or GATT panels (45 per cent) and about a third involved neither. Japan is also the most frequent target of self-initiated cases — 5 of 12 Japanese cases were self-initiated vs. only 4 of 20 EC cases. And USTR has had more success *vis-a-vis* Japan than the EC: 75 per cent of Japanese cases were at least somewhat successfully resolved vs. 40 per cent for the EC. USTR is also relatively aggressive *vis-a-vis* Korea and Taiwan, with just over half of those cases involving explicit threats and over a third being self-initiated. A slightly higher proportion of these cases was successful than of those against Japan (82 per cent).

6. SECTION 301 SINCE THE 1985 TRADE POLICY ACTION PLAN

Though only a third of section 301 cases were satisfactorily resolved from 1975 to September 1985, two-thirds have been at least partially successful since then. The difference in success rates for cases completed before and after that date is statistically significant at the one per cent level according to the chi-square test. The data show that cases completed after September 1985 were 4.5 times more likely to be successful than those concluded before that date.

A part of the increased success is probably attributable to a decrease in the proportion of cases targeting the EC, and an increase in the number of cases involving countries against whom the United States had the most success even before 1985: Japan, Korea, and Taiwan. However, the proportion of successful cases involving those countries also rose while the success rate *vis-a-vis* the EC rose more than for any other category.

The variable that showed the most change, however, was the public threat category, apparently reflecting the increased aggressiveness with which USTR used section 301 after September 1985. The success rate for cases involving public threats increased from zero to better than 70 per cent. The increase in success holds even when self-initiated cases are excluded, with over 50 per cent of privately-initiated petitions being brought to satisfactory conclusion vs. the 34 per cent of the earlier period (no cases were self-initiated prior to 1985). Among the cases that were not self-initiated, the proportion involving countries with a trade linkage greater than 30 per cent was about the same in both periods, but the proportion of cases involving public and explicit threats and the use of GATT panels increased substantially.

To test whether the Trade Policy Action Plan and 1988 Trade Act added anything to the effectiveness of section 301 that is not captured by the other independent variables, two more dummy variables were created. The first, TPAP, is equal to 1 if a case was concluded before September 1985 and 0 otherwise; a negative coefficient is expected for this variable. TBILL is equal to 1 if the case was initiated after the signing of the congressional trade bill in August 1988, and 0 otherwise; a positive coefficient is expected.[20] As shown in Table 5, the coefficient for TPAP is negative as expected and is significant at the five per cent level. The coefficient for TBILL is insignificant, but also unexpectedly negative.[21] The chi-square results also find that TPAP is significant, but that TBILL is not. Although it is still early to make definitive interpretations, these results suggest that, contrary to conventional wisdom, the 1988 Trade Act did not significantly improve the chances of achieving a successful outcome, relative to changes in the use of 301 adopted in 1985.

7. CONCLUSIONS

There is no evidence that the aggressive use of section 301, *as it has been employed to date*, has increased the risk of costly trade wars.[22] The United States has imposed retaliation in only eight cases: in four, the sanctions were lifted relatively quickly and the dispute resolved; in the other four, no resolution was achieved and the sanctions remain in place. Only one US retaliation has precipitated a counter-retaliation (in the dispute over EC preferences for citrus imports from the Mediterranean), and that episode involved less than $50 million in trade. At the same time, the trade gains reaped by the United States as a result of section 301 actions are also relatively small. An outer bound estimate of the cumulative value of US exports involved in all 67 cases examined here is less than $15 billion, well under $1 billion annually. The estimate for successful cases is less than $7 billion overall.[23]

---

[20] The end date of cases is used for TPAP because several lengthy cases that had been initiated in the late 1970s and early 1980s were concluded only after adoption of the more aggressive strategy in the fall of 1985. In fact, three of those cases — exports to Japan of leather and nonrubber footwear, and exports of canned fruit to the EC — were resolved as part of the action plan, when President Reagan ordered USTR to prepare retaliation hit lists if the cases were not resolved before December 1 of that year. The provisions of the 1988 Trade Act applied only to cases initiated after its enactment.

[21] TBILL and TPAP are negatively correlated, and if TPAP is dropped, the sign on TBILL becomes positive but is still insignificant. TPAP is also negatively correlated with DUMMY and TOUGH and the coefficients on those variables also lose significance when TPAP is added.

[22] This assessment might have to be modified, depending on the outcome of the US-EC oilseeds dispute in the summer of 1992.

A comprehensive assessment of section 301, however, should include its impact on GATT, the international trading system more generally, and US relations with its major trading partners. If the Uruguay Round concludes successfully, with reforms in the 'old' areas of agriculture and textiles and apparel, new rules to govern trade in services, intellectual property protection, and investment, and a strengthened dispute settlement mechanism, the degree to which aggressive use of section 301 may have contributed to that outcome will have to be assessed. If the Round cannot be completed, the possibility that aggressive unilateralism on the part of the United States eroded support for the GATT must also be considered.

The outcome of the Uruguay Round is also critical to any assessment of the likely future utility of section 301. One of the strongest findings in this paper was that a successful outcome in a 301 negotiation is most likely if the barrier is transparent and the international rules are clear. With more and more trade disputes occurring over issues where the current rules are unclear or nonexistent, failure to complete the Round could lower the probability of successful outcomes in 301 negotiations and increase the risk of trade wars. Conversely, a successful Round, with improved rules and dispute settlement procedures, should reduce the need for aggressive use of section 301 while increasing its effectiveness when it is needed. Thus, both critics and proponents of section 301 share an interest in the successful completion and implementation of a Uruguay Round agreement.

REFERENCES

Bayard, T.O. (1990), 'Comment on Sykes', *Boston University International Law Journal*, **8**, 325-331.

Bayard, T.O. and K.A. Elliott (forthcoming), *Reciprocity and Retaliation: An Evaluation of Aggressive Trade Policies* (Washington: Institute for International Economics).

Bello, J. (1990), 'Section 301 of the US Trade Laws: Champion of Market Liberalization', unpublished.

Bello, J. and A.F. Holmer (1990), 'The Heart of the 1988 Trade Act: A Legislative History of the Amendments to Section 301', in J. Bhagwati and H. Patrick (eds.) *Aggressive Unilateralism: America's 301 Trade Policy and the World Trading System* (Ann Arbor: University of Michigan Press).

---

[23] In most cases, this estimate is based either on the actual value of US exports to the target of the product in question; in service and intellectual property cases, as well as a few others where reliable trade data does not exist, estimates by the petitioner of the injury suffered, or announced retaliation amounts were used. These estimates probably overstate to varying degrees the actual incremental gain in US exports from these actions and are provided only to suggest an order of magnitude of the potential trade gains generated by section 301.

Bhagwati, J. and H. Patrick (eds.), (1990), *Aggressive Unilateralism: America's 301 Trade Poli and the World Trading System* (Ann Arbor: University of Michigan Press).

Cline, W.R. (1982), 'Reciprocity: A New Approach to World Trade Policy?' Policy Analysis (Washington: Institute for International Economics).

Destler, I.M. (1992), *American Trade Politics* (Institute for International Economics), 2nd e 124-25.

Hudec, R. (1990), 'Thinking About the New Section 301: Beyond Good and Evil', in J. Bhagwa and H. Patrick (eds.) *Aggressive Unilateralism: America's 301 Trade Policy and the Wor Trading System* (Ann Arbor: University of Michigan Press).

McMillan, J. (1990), 'Strategic Bargaining and Section 301', in J. Bhagwati and H. Patrick (eds *Aggressive Unilateralism: America's 301 Trade Policy and the World Trading System* (An Arbor: University of Michigan Press).

Odell, J.S. (1992), 'International Threats and Internal Politics: Brazil, the European Communit and the United States, 1985-87' in P. Evans et al. (eds.) *International Bargaining and Domest Politics: An Interactive Approach* (Berkeley: University of California Press).

Ryan, M. (1991), 'Strategy and Compliance with Bilateral Trade Dispute Settlement Agreements USTR's Section 301 Experience in the Pacific Basin', *Michigan Journal of International Law* 14, 799-827.

Schelling, T. (1960), *The Strategy of Conflict* (Cambridge Mass: Harvard University Press).

Sykes, A.O. (1990), '"Mandatory" Retaliation for Breach of Trade Agreements: Some Thought on the Strategic Design of Section 301', *Boston University International Law Journal*, 8 301-24.

35

WHARTON REPROGRAPHICS

The Case for Free Trade

Environmentalists are wrong to fear the effects of free trade. Both causes can be advanced by imaginative solutions

by Jagdish Bhagwati

Economists are reconciled to the conflict of absolutes: that is why they invented the concept of tradeoffs. It should not surprise them, therefore, that the objective of environmental protection should at times run afoul of the goal of seeking maximum gains from trade. In fact, economists would be suspicious of any claims, such as those made by soothsaying politicians, that both causes would be only mutually beneficial. They are rightly disconcerted, however, by the passion and the ferocity, and hence often the lack of logic or facts, with which environmental groups have recently assailed both free trade and the General Agreement on Tariffs and Trade (GATT), the institution that oversees the world trading system.

The environmentalists' antipathy to trade is perhaps inevitable. Trade has been central to economic thinking since Adam Smith discovered the virtues of specialization and of the markets that naturally sustain it. Because markets do not normally exist for the pursuit of environmental protection, they must be specially created. Trade therefore suggests abstention from governmental intervention, whereas environmentalism suggests its necessity. Then again, trade is exploited and its virtues extolled by corporate and multinational interests, whereas environmental objectives are embraced typically by nonprofit organizations, which are generally wary of these interests. Trade is an ancient occupation, and its nurture is the objective of institutions crafted over many years of experience and reflection. Protection of the environment, on the other hand, is a recent preoccupation of national and international institutions that are nascent and still evolving.

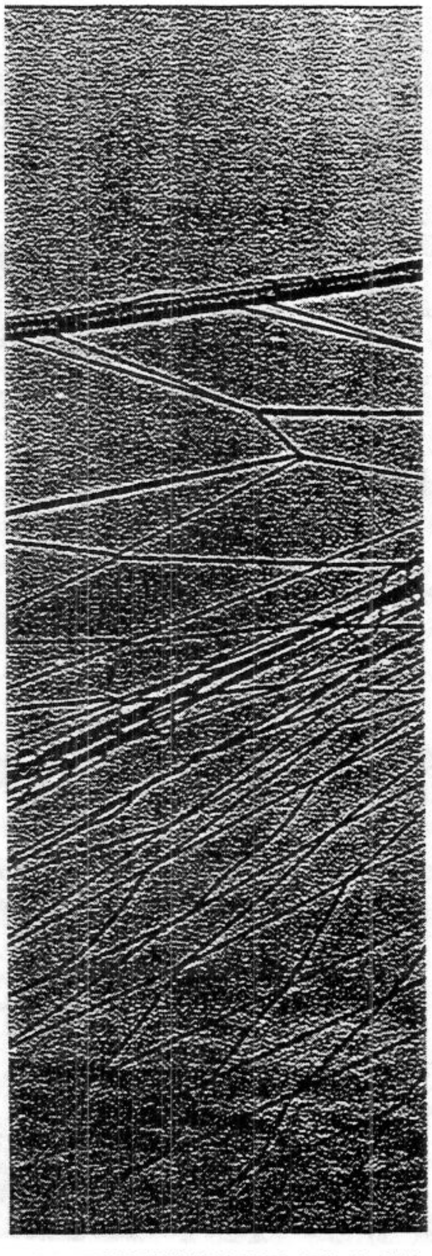
JAGDISH BHAGWATI is Arthur Lehman Professor of Economics and professor of political science at Columbia University and was Ford International Professor of Economics at the Massachusetts Institute of Technology. He has served as the economic policy adviser to the director-general of the General Agreement on Tariffs and Trade. Five volumes of his collected essays have been published by MIT Press. His most recent books are *Protectionism* (MIT Press, 1988) and *The World Trading System at Risk* (Princeton University Press, 1991). He also writes frequently for the *New York Times*, the *Wall Street Journal* and the *New Republic*.

Last year the environmentalists' hostility to trade exploded in outrage when an impartial GATT Dispute Settlement Panel ruled in favor of Mexico and free trade and against the U.S. and the welfare of the dolphin. The U.S. had placed an embargo on the import of Mexican tuna on the grounds that the fish had been caught in purse-seine nets, which kill dolphins cruelly and in greater numbers than U.S. law permits. The GATT panel ruled, in effect, that the U.S. could not suspend Mexico's trading rights by proscribing unilaterally the methods by which that country harvested tuna.

This decision spurred the conservationists' subsequent campaigns against free trade and GATT. GATT has no shortage of detractors, of course. In fact, some of its recent critics have feared its impotence and declared it "dead," referring to it as the General Agreement to Talk and Talk. But the environmentalist attacks, which presume instead GATT's omnipotence, are something else again.

An advertisement by a coalition of environmental groups in the *New York Times* on April 20, 1992, set a new standard for alarmist, even scurrilous, writing, calculated to appeal to one's instincts rather than one's intellect. It talks of "faceless GATT bureaucrats" mounting a "sneak attack on democracy." This veiled reference to Pearl Harbor provides an example of a common tactic in trade controversy: Japan-bashing. The innuendos have continued unabated and are manifest in the endless battles in Congress over the supplemental environmental accords for the North American Free Trade Agreement (NAFTA). The hostility is also intruding on the conclusion of the Uruguay Round of GATT talks, now in their seventh year, with the environmentalists opposing the establishment of the new Multilateral Trade Organization, which is meant to provide effective discipline and a necessary institutional structure for GATT.

It is surely tragic that the proponents of two of the great causes of the 1990s, trade and the environment, should be



DOLPHIN VERSUS FREE TRADE: the U.S. outlaws fishing methods that result in the death of dolphins such as this one, locked in combat. The conflict is largely gratuitous. There are at times philosophical differences between the two that cannot be reconciled, as when some environmentalists assert nature's autonomy, whereas most economists see nature as a handmaiden to humankind. For the most part, however, the differences derive from misconceptions. It is necessary to dissect and dismiss the more egregious of these fallacies before addressing the genuine problems.

The fear is widespread among environmentalists that free trade increases economic growth and that growth harms the environment. That fear is misplaced. Growth enables governments to tax and to raise resources for a variety of objectives, including the abatement of pollution and the general protection of the environment. Without such revenues, little can be achieved, no matter how pure one's motives may be.

How do societies actually spend these additional revenues? It depends on how getting rich affects the desire for a better environment. Rich countries today have more groups worrying about environmental causes than do poor countries. Efficient policies, such as freer trade, should generally help environmentalism, not harm it.

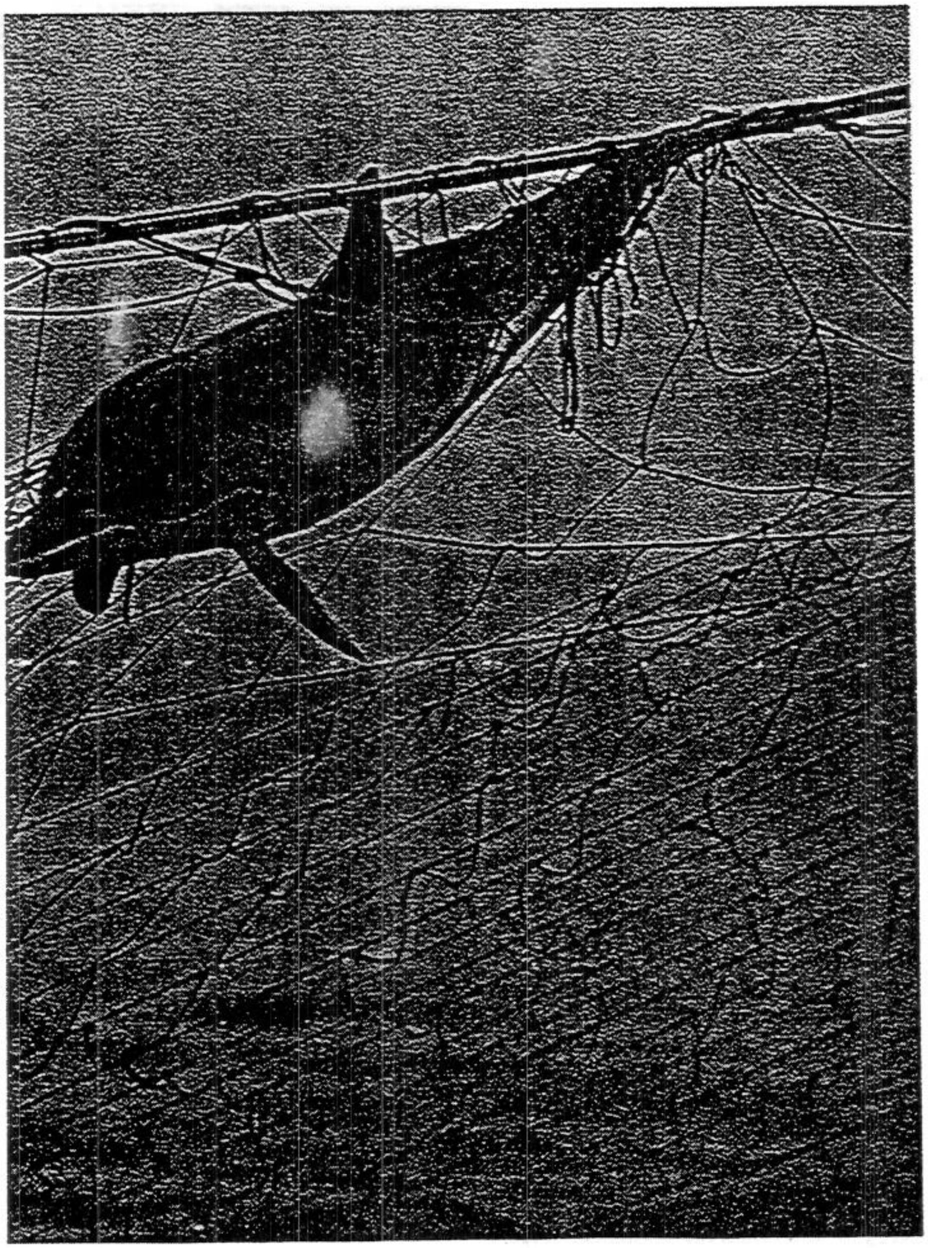
If one wants to predict what growth will do to the environment, however, one must also consider how it will affect the production of pollution. Growth affects not only the demand for a good environment but also the supply of the pollution associated with growth. The net effect on the environment will therefore depend on the kind of economic growth. Gene M. Grossman and Alan B. Krueger of Princeton University found that in cities around the world sulfur dioxide pollution fell as per capita income rose. The only exception was in countries whose per capita incomes fell below $5,000. In short, environmentalists are in error when they fear that trade, through growth, will necessarily increase pollution.

Economic effects besides those attributable to rising incomes also help to protect the environment. For example, freer trade enables pollution-fighting technologies available elsewhere to be imported. Thus, trade in low-sulfur-content coal will enable the users of local high-sulfur-content coal to shift from the latter to the former.

Free trade can also lead to better environmental outcomes from a shift in the composition of production. An excellent example is provided by Robert C. Feenstra of the University of California at Davis. He has shown how the imposition of restraints on Japanese automobile exports to the U.S. during the 1980s shifted the composition of those exports from small to large cars, as the Japanese attempted to increase their revenues without increasing the number of units they sold. Yet the large cars were fuel inefficient. Thus, protective efforts by the U.S. effectively increased the average amount of pollution produced by imported cars, making it more likely that pollution from cars would increase rather than diminish in the U.S.

Although these erroneous objections to free trade are readily dismissed (but not so easily eliminated from public discourse), there are genuine conflicts between trade and the environment. To understand and solve them, economists draw a distinction between two kinds of environmental problems: those that are intrinsically domestic and those that are intrinsically transnational.

Should Brazil pollute a lake lying wholly within its borders, the problem would be intrinsically domestic. Should it pollute a river that flows into Argentina, the matter would take on an intrinsically transnational character. Perhaps the most important examples of transnational pollution are acid rain, created when sulfur dioxide emissions in one country precipitate into rain in another, and greenhouse gases, such as carbon dioxide, which contribute to global warming wherever they are emitted.



ensnared off the U.S. Atlantic coast. But when the U.S. attempted to apply its standard to Mexico by imposing an embargo on tuna imported from that country, an international tribunal rejected the policy last year as an illegal restriction of trade.

Why do intrinsically domestic environmental questions create international concern? The main reason is the belief that diversity in environmental standards may affect competitiveness. Businesses and labor unions worry that their rivals in other countries may gain an edge if their governments impose lower standards of environmental protection. They decry such differences as unfair. To level the playing field, these lobbies insist that foreign countries raise their standards up to domestic ones. In turn, environmental groups worry that if such "harmonization up" is not undertaken prior to freeing trade, pressures from uncompetitive businesses at home will force down domestic standards, reversing their hard-won victories. Finally, there is the fear, dramatized by H. Ross Perot in his criticisms of NAFTA, that factories will relocate to the countries whose environmental standards are lowest.

But if the competitiveness issue makes the environmentalists, the businesses and the unions into allies, the environmentalists are on their own in other ways. Two problem areas can be distinguished. First, some environmentalists are keen to impose their own ethical preferences on others, using trade sanctions to induce or coerce acceptance of such preferences. For instance, tuna fishing with purse-seine nets that kill dolphins is opposed by U.S. environmental groups, which consequently favor restraints on the importation of such tuna from Mexico and elsewhere. Second, other environmentalists fear that the rules of free trade, as embodied in GATT and strengthened in the Uruguay Round, will constrain their freedom to pursue even purely domestic environmental objectives, with GATT tribunals outlawing disputed regulation.

Environmentalists have cause for concern. Not all concerns are legitimate, however, and not all the solutions to legitimate concerns are sensible. Worry over competitiveness has thus led to the illegitimate demand that environmental standards abroad be treated as "social dumping." Offending countries are regarded as unfairly subsidizing their exporters through lax environmental requirements. Such implicit subsidies, the reasoning continues, ought to be offset by import duties.

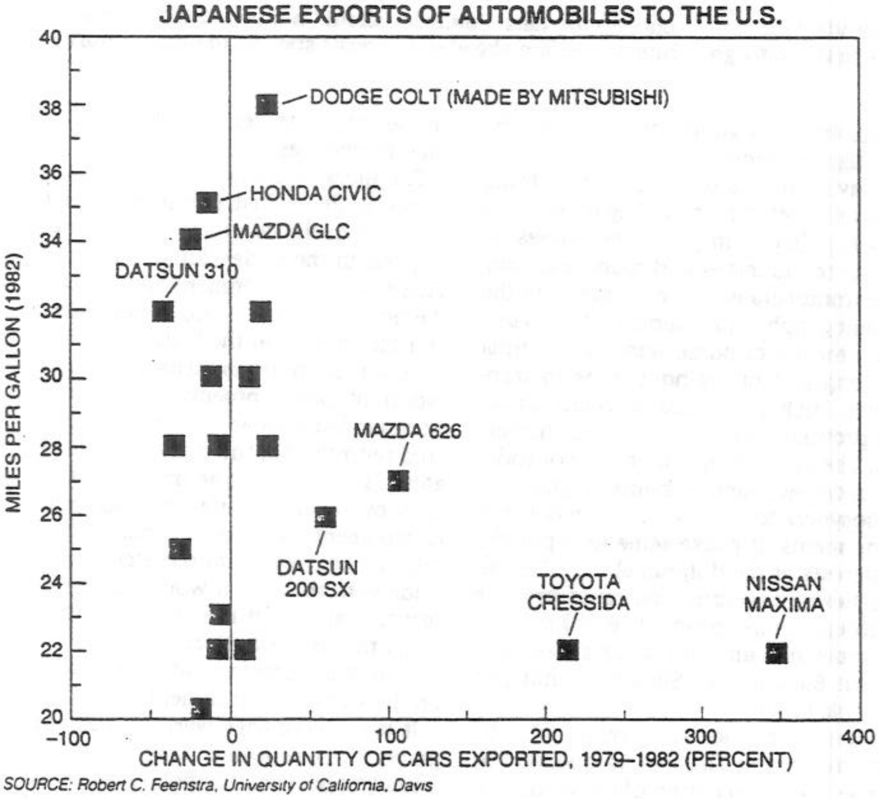
Yet international differences in environmental standards are perfectly natural. Even if two countries share the same environmental objectives, the *specific* pollutions they would attack, and hence the industries they would hinder, will generally not be identical. Mexico has a greater social incentive than does the U.S. to spend an extra dollar preventing dysentery rather than reducing lead in gasoline.

Equally, a certain environmental good might be valued more highly by a poor country than by a rich one. Contrast, for instance, the value assigned to a lake with the cost of cleaning up effluents discharged into it by a pharmaceutical company. In India such a lake's water might be drunk by a malnourished population whose mortality would increase sharply with the rise in pollution. In the U.S. the water might be consumed by few people, all of whom have the means to protect themselves with privately purchased water filters. In this example, India would be the more likely to prefer clean water to the pharmaceutical company's profits.

The consequences of differing standards are clear: each country will have less of the industry whose pollution it fears relatively more than other countries do. Indeed, even if there were no international trade, we would be shrinking industries whose pollution we deter. This result follows from the policy of forcing polluters of all stripes to pay for the harm they cause. To object, then, to the effects our negative valuation of pollution have on a given industry is to be in contradiction: we would be refusing to face the consequences of our environmental preferences.

Nevertheless, there is sentiment for enacting legislation against social dumping. Senator David L. Boren of Oklahoma, the proponent of the International Pollution Deterrence Act of 1991, demanded import duties on the grounds that "some U.S. manufacturers, such as the U.S. carbon and steel alloy industry, spend as much as 250 percent more on environmental controls as a percentage of gross domestic product than do other countries.... I see the unfair advantage enjoyed by other nations exploiting the environment and public health for economic gain when I look at many industries important to my own state." Similarly, Vice President Al Gore wrote in *Earth in the Balance: Ecology and the Human Spirit* that "just as government subsidies of a particular industry are sometimes considered unfair under the trade laws, weak and ineffectual enforcement of pollution control measures should also be included in the definition of unfair trading practices."

These demands betray lack of economic logic, and they ignore political reality as well. Remember that the so-called subsidy to foreign producers through lower standards is not given but only implied. According to Senator



SOURCE: Robert C. Feenstra, University of California, Davis

PERVERSE CONSEQUENCES for the environment may result from trade restrictions. This graph shows Japanese car exports to the U.S. before and after Japan's acquiescence in voluntary export restraints. Sales of small, fuel-efficient models declined, whereas those of the larger "gas guzzlers" soared.

EMPLOYMENT IN MEXICAN TUNA FISHERY may offset the saving of dolphins that would result were the industry to forgo purse-seine nets. Countries should not be faulted for placing human welfare ahead of concerns specific to other cultures.

Boren, the subsidy would be calculated as "the cost that would have to be incurred by the manufacturer or producer of the foreign articles of merchandise to comply with environmental standards imposed on U.S. producers of the same class of merchandise." Anyone familiar with the way dumping calculations are made knows that the Environmental Protection Agency could come up with virtually any estimates it cared to produce. Cynical politics would inevitably dictate the calculations.

Still, there may be political good sense in assuaging environmentalists' concerns about the relocation of factories to countries with lower standards. The governments of higher-standards countries could do so without encumbering free trade by insisting that their businesses accede to the higher standards when they go abroad. Such a policy lies entirely within the jurisdictional powers of a higher-standards country. Moreover, the governments of lower-standards countries would be most unlikely to object to such an act of good citizenship by the foreign investors.

Environmentalists oppose free trade for yet another reason: they wish to use trade policy to impose their values on other communities and countries. Many environmentalists want to suspend the trading rights of countries that sanction the use of purse-seine nets in tuna fishing and of leg-hold traps in trapping. Such punishments seem an inappropriate use of state power, however. The values in question are not widely accepted, such as human rights, but idiosyncratic. One wonders when the opponents of purse-seine nets put the interests of the dolphin ahead of those of Mexico's people, who could prosper through more productive fishing. To borrow the campaign manifesto of President Bill Clinton: Should we not put people first?

Moreover, once such values intrude on free trade, the way is opened for an endless succession of demands. Environmentalists favor dolphins; Indians have their sacred cows. Animal-rights activists, who do not prefer one species over another, will object to our slaughterhouses.

The moral militancy of environmentalists in the industrialized world has begun to disillusion their closest counterparts in the undeveloped countries. These local environmentalists accuse the rich countries of "eco-imperialism," and they deny that the Western nations have a monopoly on virtue. The most radical of today's proenvironment magazines in India, *Down to Earth*, editorialized recently: "In the current world reality trade is used as an instrument entirely by Northern countries to discipline environmentally errant nations. Surely, if India or Kenya were to threaten to stop trade with the U.S., it would hardly affect the latter. But the fact of the matter is that it is the Northern countries that have the greatest [adverse] impact on the world's environment."

If many countries were to play this game, then repeated suspensions of trading rights would begin to undermine the openness of the trading system and the predictability and stability of international markets. Some environmentalists assert that each country should be free to insist on the production methods of its trading partners. Yet these environmentalists ignore the certain consequence of their policy: a Pandora's box of protectionism would open up. Rarely are production methods in an industry identical in different countries.

There are certainly better ways to indulge the environmentalists' propensity to export their ethical preferences. The U.S. environmental organizations can lobby in Mexico to persuade its government to adopt their views. Private boycotts can also be undertaken. In fact, boycotts can carry much clout in rich countries with big markets, on which the targeted poor countries often depend. The frequent and enormously expensive advertisements by environmental groups against GATT show also that their resources far exceed those of the cash-strapped countries whose policies they oppose.

Cost-benefit analysis leads one to conclude that unilateral governmental suspension of others' trading rights is not an appropriate way to promote one's lesser ethical preferences. Such sanctions can, on the other hand, appropriately be invoked multilaterally to defend universal moral values. In such cases—as in the censure of apartheid, as practiced until recently in South Africa—it is possible to secure widespread agreement for sanctions. With a large majority converted to the cause, GATT's waiver procedure can be used to suspend the offending country's trading rights.

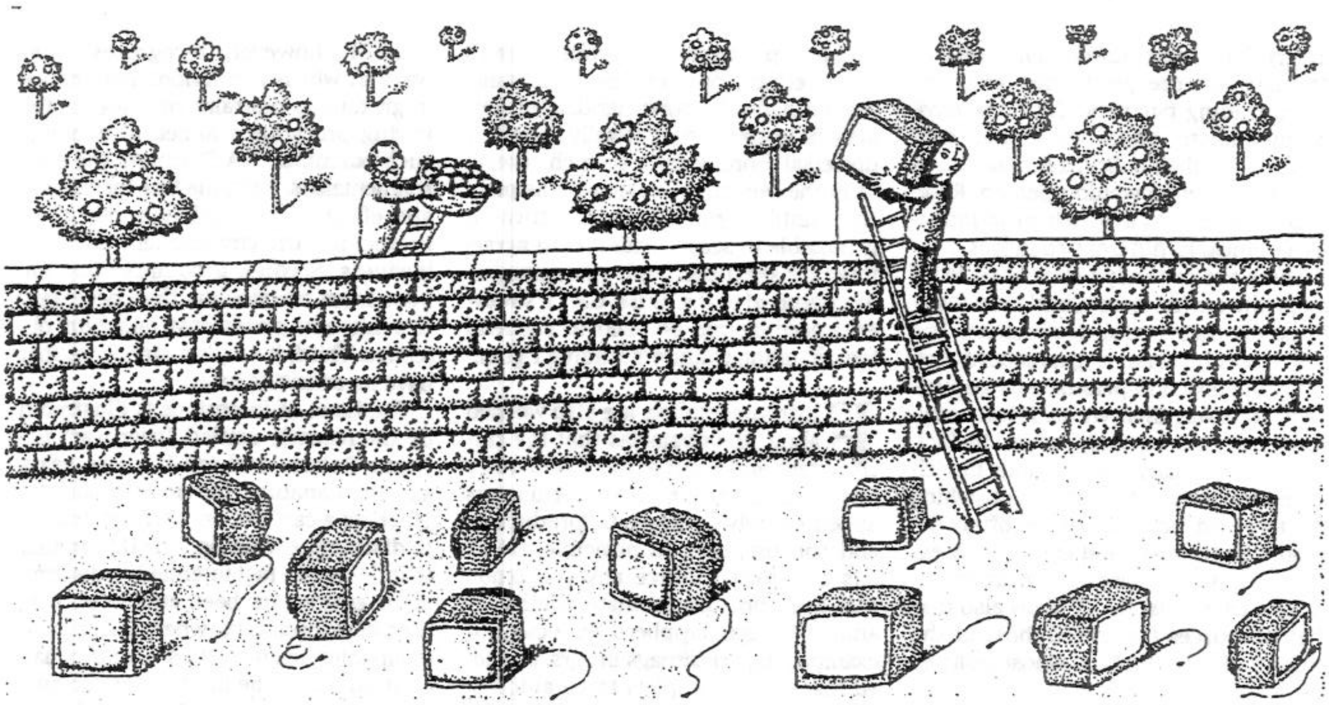
Environmentalists are also worried about the obstacles that the current and prospective GATT rules pose for environmental regulations aimed entirely at domestic production and consumption. In principle, GATT lets a country enforce any regulation that does not discriminate against or among foreign suppliers. One can, for example, require airbags in cars, provided that the rule applies to all automobile makers. GATT even permits rules that discriminate against trade for the purpose of safety and health.

GATT, however, recognizes three ways in which regulations may be set in gratuitous restraint of trade; in following procedures aimed at avoiding such outcomes, GATT upsets the environmentalists. First, the true intention—and effect—of a regulation may be to protect not the environment but local business. Second, a country may impose more restrictions than necessary to achieve its stated environmental objective. Third, it may set standards that have no scientific basis.

The issue of intentions is illustrated by the recently settled "beer war" between Ontario and the U.S. Five years ago the Canadian province imposed a 10-cents-a-can tax on beer, ostensibly to discourage littering. The U.S. argued that the law in fact intended to discriminate against its beer suppliers, who used aluminum cans, whereas local beer companies used bottles. Ontario had omitted to tax the use of cans for juices and soups, a step that would have affected Ontario producers.

The second problem is generally

PURE DRINKING WATER is essential for Mexican villagers, who wait in line to collect it rather than risk contracting cholera from local sources. The relative value of environmental benefits varies in different countries: Mexico can better improve public health by concentrating its resources on the purification of water than by reducing the lead in gasoline.



BENEFITS OF TRADE flow from the economies achieved when countries specialize in enterprises in which they enjoy comparative advantage. Such specialization will proceed better when all sides trust in the stability of the trading regime.

tougher because it is impossible to find alternative restrictions that accomplish exactly the same environmental results as the original policy at lower cost. An adjudicating panel is then forced to evaluate, implicitly or explicitly, the trade-offs between the cost in trade disruption and the cost in lesser fulfillment of the environmental objective. It is therefore likely that environmentalists and trade experts will differ on which weights the panel should assign to these divergent interests.

Environmentalists tend to be fearful about the use of scientific tests to determine whether trade in a product can be proscribed. The need to prove one's case is always an unwelcome burden to those who have the political power to take unilateral action. Yet the trade experts have the better of the argument. Imagine that U.S. growers sprayed apples with the pesticide Alar, whereas European growers did not, and that European consumers began to agitate against Alar as harmful. Should the European Community be allowed to end the importation of the U.S. apples without meeting *some* scientific test of its health concerns? Admittedly, even hard science is often not hard enough—different studies may reach different conclusions. But without the restraining hand of science, the itch to indulge one's fears—and to play on the fears of others—would be irresistible.

In all cases, the moderate environmentalists would like to see GATT adopt more transparent procedures for adjudicating disputes. They also desire greater legal standing to file briefs when environmental regulations are at issue. These goals seem both reasonable and feasible.

Not all environmental problems are local; some are truly global, such as the greenhouse effect and the depletion of the stratospheric ozone. They raise more issues that require cooperative, multilateral solutions. Such solutions must be both efficient and equitable. Still, it is easy to see that rich countries might use their economic power to reach protocols that maximize efficiency at the expense of poorer countries.

For instance, imagine that the drafters of a protocol were to ask Brazil to refrain from cutting down its rain forests while allowing industrialized countries to continue emitting carbon dioxide. They might justify this request on the grounds that it costs Brazil less to keep a tree alive, absorbing a unit of carbon dioxide every year, than it would cost the U.S. or Germany to save a unit by burning less oil. Such a trade-off would indeed be economically efficient. Yet if Brazil, a poorer country, were then left with the bill, the solution would assuredly be inequitable.

Before any group of countries imposes trade sanctions on a country that has not joined a multilateral protocol, it would be important to judge whether the protocol is indeed fair. Nonmembers targeted for trade sanctions should have the right to get an impartial hearing of their objections, requiring the strong to defend their actions even when they appear to be entirely virtuous.

The simultaneous pursuit of the two causes of free trade and a protected environment often raises problems, to be sure. But none of these conflicts is beyond resolution with goodwill and by imaginative institutional innovation. The aversion to free trade and GATT that many environmentalists display is unfounded, and it is time for them to shed it. Their admirable moral passion and certain intellectual vigor are better devoted to building bridges between the causes of trade and the environment.

FURTHER READING

AMERICAN RULES, MEXICAN JOBS. Jagdish Bhagwati in *New York Times*, Section A, page 21, col. 1; March 24, 1993.

"CIRCUMVENTING" DEMOCRACY: THE POLITICAL MORALITY OF TRADE NEGOTIATIONS. Robert E. Hudec in *New York University Journal of International Law and Politics*, Vol. 25, No. 2, pages 401–412; September/October 1993.

ENVIRONMENTAL IMPACTS OF A NORTH AMERICAN FREE TRADE AGREEMENT. Gene M. Grossman and Alan B. Krueger in *The Mexico-U.S. Free Trade Agreement*. Edited by Peter M. Garber. MIT Press, 1993.

TRADE AND ENVIRONMENT: DOES ENVIRONMENTAL DIVERSITY DETRACT FROM THE CASE FOR FREE TRADE? Jagdish Bhagwati and T. N. Srinivasan. Mimeograph. Yale University, 1993.

36

WHARTON REPROGRAPHICS

The Perils of Free Trade

Economists routinely ignore its hidden costs to the environment and the community

by Herman E. Daly

No policy prescription commands greater consensus among economists than that of free trade based on international specialization according to comparative advantage. Free trade has long been presumed good unless proved otherwise. That presumption is the cornerstone of the existing General Agreement on Tariffs and Trade (GATT) and the proposed North American Free Trade Agreement (NAFTA). The proposals in the Uruguay Round of negotiations strengthen GATT's basic commitment to free trade and economic globalization.

Yet that presumption should be reversed. The default position should favor domestic production for domestic markets. When convenient, balanced international trade should be used, but it should not be allowed to govern a country's affairs at the risk of environmental and social disaster. The domestic economy should be the dog and international trade its tail. GATT seeks to tie all the dogs' tails together so tightly that the international knot would wag the separate national dogs.

The wiser course was well expressed in the overlooked words of John Maynard Keynes: "I sympathize, therefore, with those who would minimize, rather than those who would maximize, economic entanglement between nations. Ideas, knowledge, art, hospitality, travel—these are the things which should of their nature be international. But let goods be homespun whenever it is reasonably and conveniently possible; and, above all, let finance be primarily national." Contrary to Keynes, the defenders of the proposed Uruguay Round of changes to GATT not only want to downplay "homespun goods," they also want finance and all other services to become primarily international.

HERMAN E. DALY is senior economist in the environment department of the World Bank in Washington, D.C. Before joining the bank in 1988, he was alumni professor of economics at Louisiana State University. He holds a B.A. from Rice University and a Ph.D. from Vanderbilt University. Daly has taught in Brazil as a Ford Foundation Visiting Professor and as a Senior Fulbright Scholar. He has also served as a research associate at Yale University and as a visiting fellow at the Australian National University. Co-founder and associate editor of *Ecological Economics*, Daly has written several books, including *Steady-State Economics*. The views expressed here by Daly should not be attributed to the World Bank.

Economists and environmentalists are sometimes represented as being, respectively, for and against free trade, but that polarization does the argument a disservice. Rather the real debate is over what kinds of regulations are to be instituted and what goals are legitimate. The free traders seek to maximize profits and production without regard for considerations that represent hidden social and environmental costs. They argue that when growth has made people wealthy enough, they will have the funds to clean up the damage done by growth. Conversely, environmentalists and some economists, myself among them, suspect that growth is increasing environmental costs faster than benefits from production—thereby making us poorer, not richer.

A more accurate name than the persuasive label "free trade"—because who can be opposed to freedom?—is "deregulated international commerce." Deregulation is not always a good policy: recall the recent experience of the U.S. with the deregulation of the savings and loan institutions. As one who formerly taught the doctrine of free trade to college students, I have some sympathy for the free traders' view. Nevertheless, my major concern about my profession today is that our disciplinary preference for logically beautiful results over factually grounded policies has reached such fanatical proportions that we economists have become dangerous to the earth and its inhabitants.

The free trade position is grounded in the logic of comparative advantage, first explicitly formulated by the early 19th-century British economist David Ricardo. He observed that countries with different technologies, customs and resources will incur different costs when they make the same products. One country may find it comparatively less costly to mine coal than to grow wheat, but in another country the opposite may be true. If nations specialize in the

POLLUTING is one way in which industries can "externalize" some of the costs associated with production. Industries have profit incentives to produce goods

products for which they have a comparative advantage and trade freely to obtain others, everyone benefits.

The problem is not the logic of this argument. It is the relevance of Ricardo's critical but often forgotten assumption that factors of production (especially capital) are internationally immobile. In today's world, where billions of dollars can be transferred between nations at the speed of light, that essential condition is not met. Moreover, free traders encourage such foreign investment as a development strategy. In short, the free traders are using an argument that hinges on the impermability of national boundaries to capital to support a policy aimed at making those same boundaries increasingly permeable to both capital and goods!

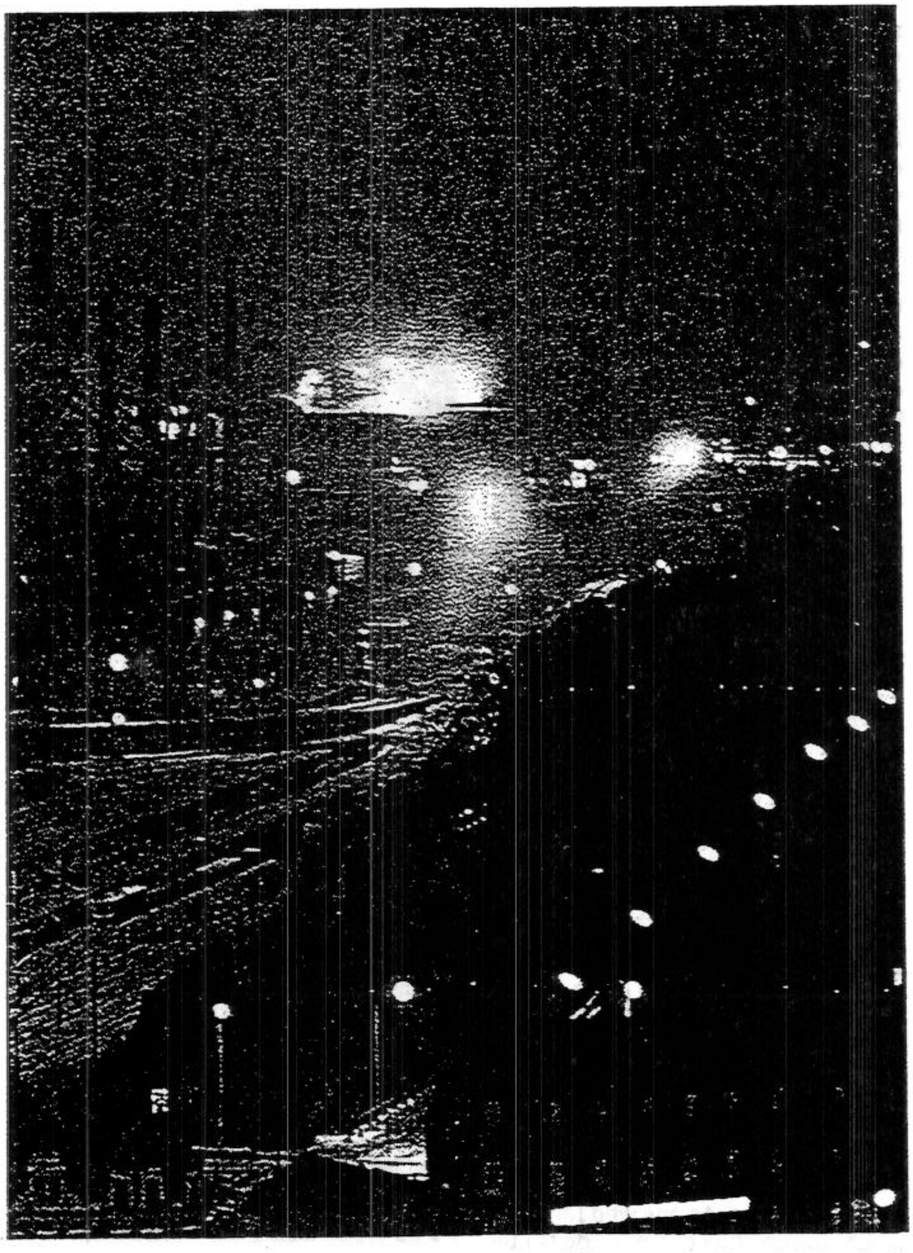
That fact alone invalidates the assumption that international trade will inevitably benefit all its partners. Furthermore, for trade to be mutually beneficial, the gains must not be offset by higher liabilities. After specialization, nations are no longer free *not* to trade, and that loss of independence can be a liability. Also, the cost of transporting goods internationally must not cancel out the profits. Transport costs are energy intensive. Today, however, the cost of energy is frequently subsidized by governments through investment tax credits, federally subsidized research and military expenditures that ensure access to petroleum. The environmental costs of fossil-fuel burning also do not factor into the price of gasoline. To the extent that energy is subsidized, then, so too is trade. The full cost of energy, stripped of these obscuring subsidies, would therefore reduce the initial gains from long-distance trade, whether international or interregional.

Free trade can also introduce new inefficiencies. Contrary to the implications of comparative advantage, more than half of all international trade involves the simultaneous import and export of essentially the same goods. For example, Americans import Danish sugar cookies, and Danes import American sugar cookies. Exchanging recipes would surely be more efficient. It would also be more in accord with Keynes's dictum that knowledge should be international and goods homespun (or in this case, homebaked).

Another important but seldom mentioned corollary of specialization is a reduction in the range of occupational choices. Uruguay has a clear comparative advantage in raising cattle and sheep. If it adhered strictly to the rule of specialization and trade, it would afford its citizens only the choice of being either cowboys or shepherds. Yet Uruguayans feel a need for their own legal, financial, medical, insurance and educational services, in addition to basic agriculture and industry. That diversity entails some loss of efficiency, but it is necessary for community and nationhood.

Uruguay is enriched by having a symphony orchestra of its own, even though it would be cost-effective to import better symphony concerts in exchange for wool, mutton, beef and leather. Individuals, too, must count the broader range of choices as a welfare gain: even those who are cowboys and shepherds are surely enriched by contact with countrymen who are not *vaqueros* or *pastores*. My point is that the community dimension of welfare is completely overlooked in the simplistic argument that if specialization and trade increase the per capita availability of commodities, they must be good.

Let us assume that even after those liabilities are subtracted from the gross returns on trade, positive net gains still exist. They must still offset deeper, more fundamental problems. The arguments for free trade run afoul of the three basic goals of all economic policies: the efficient *allocation* of resources, the fair *distribution* of resources and the maintenance of a sustainable *scale* of resource use. The first two are tradi-



in countries with permissive pollution, health and labor standards and then to sell the goods elsewhere. Yet that competitive pressure can drive down higher standards. Tariffs that eliminate these unfair advantages are therefore essential for protecting the global efficiency of resource use.

tional goals of neoclassical economics. The third has only recently been recognized and is associated with the viewpoint of ecological, or steady-state, economics. It means that the input of raw materials and energy to an economy and the output of waste materials and heat must be within the regenerative and absorptive capacities of the ecosystem.

In neoclassical economics the efficient allocation of resources depends on the counting and internalization of all costs. Costs are internalized if they are directly paid by those entities responsible for them—as when, for example, a manufacturer pays for the disposal of its factory wastes and raises its prices to cover that expense. Costs are externalized if they are paid by someone else—as when the public suffers extra disease, stench and nuisance from uncollected wastes. Counting all costs is the very basis of efficiency.

Economists rightly urge nations to follow a domestic program of internalizing costs into prices. They also wrongly urge nations to trade freely with other countries that do not internalize their costs (and consequently have lower prices). If a nation tries to follow both those policies, the conflict is clear: free competition between different cost-internalizing regimes is utterly unfair.

International trade increases competition, and competition reduces costs. But competition can reduce costs in two ways: by increasing efficiency or by lowering standards. A firm can save money by lowering its standards for pollution control, worker safety, wages, health care and so on—all choices that externalize some of its costs. Profit-maximizing firms in competition always have an incentive to externalize their costs to the degree that they can get away with it.

For precisely that reason, nations maintain large legal, administrative and auditing structures that bar reductions in the social and environmental standards of domestic industries. There are no analogous international bodies of law and administration; there are only national laws, which differ widely. Consequently, free international trade encourages industries to shift their production activities to the countries that have the lowest standards of cost internalization—hardly a move toward global efficiency.

Attaining cheapness by ignoring real costs is a sin against efficiency. Even GATT recognizes that requiring citizens of one country to compete against foreign prison labor would be carrying standards-lowering competition too far. GATT therefore allows the imposition of restrictions on such trade. Yet it makes no similar exception for child labor, for uninsured risky labor or for subsistence-wage labor.

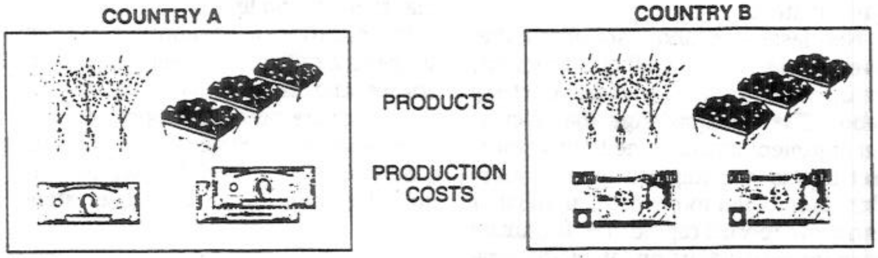
The most practical solution is to permit nations that internalize costs to levy compensating tariffs on trade with nations that do not. "Protectionism"—shielding an inefficient industry against more efficient foreign competitors—is a dirty word among economists. That is very different, however, from protecting an efficient national policy of full-cost pricing from standards-lowering international competition.

Such tariffs are also not without precedent. Free traders generally praise the fairness of "antidumping" tariffs that discourage countries from trading in goods at prices below their production costs. The only real difference is the decision to include the costs of environmental damage and community welfare in that reckoning.

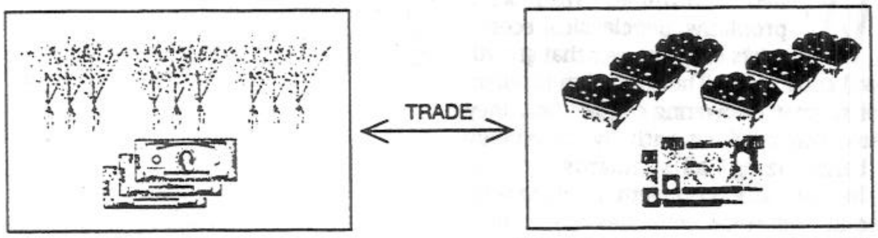
This tariff policy does not imply the imposition of one country's environmental preferences or moral judgments on another country. Each country should set the rules of cost internalization in its own market. Whoever sells in a nation's market should play by that nation's rules or pay a tariff sufficient to remove the competitive advantage of lower standards. For instance, under the Marine Mammal Protection Act, all tuna sold in the U.S. (whether by U.S. or Mexican fishermen) must count the cost of limiting the kill of dolphin associated with catching tuna. Tuna sold in the Mexican market (whether by U.S. or Mexican fishermen) need not include that cost. No standards are being imposed through "environmental imperialism"; paying the costs of a nation's environmental standards is merely the price of admission to its market.

Indeed, free trade could be accused of reverse environmental imperialism. When firms produce under the most permissive standards and sell their products elsewhere without penalty, they press on countries with higher stan-

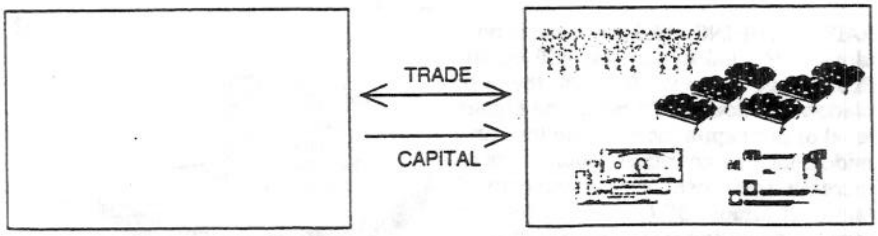
How Comparative Advantage Works



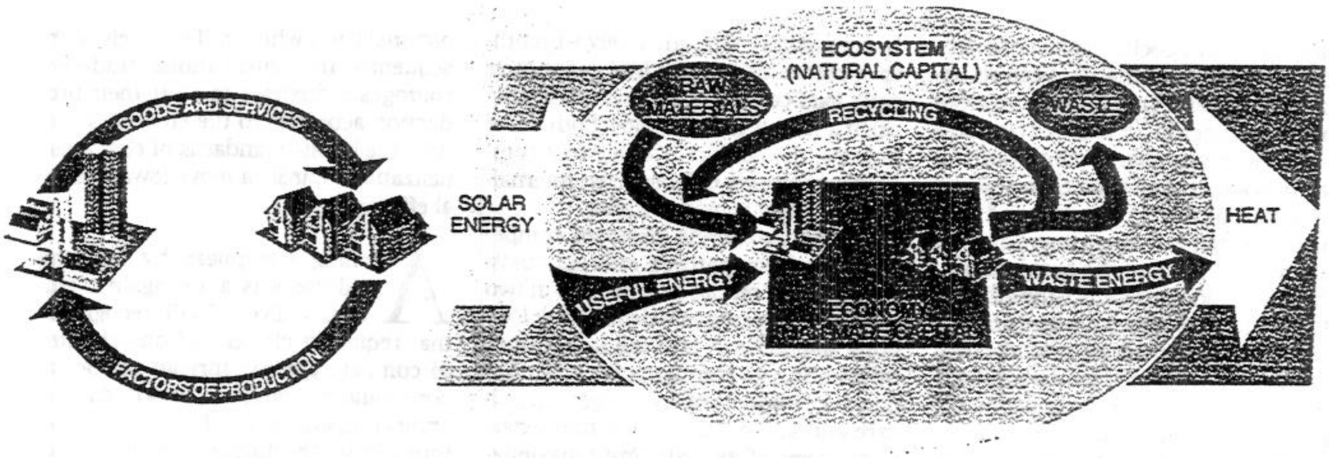
When there is no international trade, each country's production is limited entirely by its own capital and resources. Some products are comparatively less expensive to produce than others on a per unit basis.



When there is free trade, countries can specialize based on comparative advantage. All of a country's capital can be invested in making one product. Absolute cost differences between the countries do not matter. The hidden assumption is that capital cannot cross borders.



If capital is also mobile, capital can follow absolute advantage rather than comparative advantage. As in this example, one country may end up producing everything if it has lower absolute costs.



DIFFERENT VIEWS OF ECONOMIES distinguish neoclassical and steady-state economics. Neoclassical economics pictures the economy as an isolated system (*left*) in which exchange value circulates between industries and households. Neither matter nor energy enters or leaves the system, so the economy can be of any size. In the steady-state view (*right*) the economy is only one component of a larger ecosystem in which materials are transformed and energy is converted to heat. As the economy grows larger, its behavior must conform more closely to that of the total ecosystem.

dards to lower them. In effect, unrestricted trade imposes lower standards.

Unrestricted international trade also raises problems of resource distribution. In the world of comparative advantage described by Ricardo, a nation's capital stays at home, and only goods are traded. If firms are free to relocate their capital internationally to wherever their production costs would be lowest, then the favored countries have not merely a comparative advantage but an absolute advantage. Capital will drain out of one country and into another, perhaps making what H. Ross Perot called "a giant sucking sound" as jobs and wealth move with it. This specialization will increase world production, but without any assurance that all the participating countries will benefit.

When capital flows abroad, the opportunity for new domestic employment diminishes, which drives down the price for domestic labor. Even if free trade and capital mobility raise wages in low-wage countries (and that tendency is thwarted by overpopulation and rapid population growth), they do so at the expense of labor in the high-wage countries. They thereby increase income inequality there. Most citizens are wage earners. In the U.S., 80 percent of the labor force is classified as "nonsupervisory employees." Their real wages have fallen 17 percent between 1973 and 1990, in significant part because of trade liberalization.

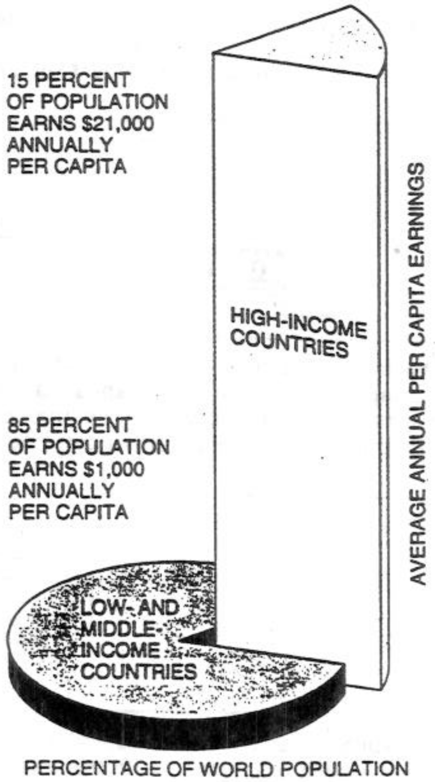
Nor does labor in low-wage countries necessarily gain from free trade. It is likely that NAFTA will ruin Mexican peasants when "inexpensive" U.S. corn (subsidized by depleting topsoil, aquifers, oil wells and the federal treasury) can be freely imported. Displaced peasants will bid down wages. Their land will be bought cheaply by agribusinesses to produce fancy vegetables and cut flowers for the U.S. market. Ironically, Mexico helps to keep U.S. corn "inexpensive" by exporting its own vanishing reserves of oil and genetic crop variants, which the U.S. needs to sustain its corn monoculture.

Neoclassical economists admit that overpopulation can spill over from one country to another in the form of cheap labor. They acknowledge that fact as an argument against free immigration. Yet capital can migrate toward abundant labor even more easily than labor can move toward capital. The legitimate case for restrictions on labor immigration is therefore easily extended to restrictions on capital emigration.

When confronted with such problems, neoclassical economists often answer that growth will solve them. The allocation problem of standards-lowering competition, they say, will be dealt with by universally "harmonizing" all standards upward. The distribution problem of falling wages in high-wage countries would only be temporary; the economists believe that growth will eventually raise wages worldwide to the former high-wage level and beyond.

Yet the goal of a sustainable scale of total resource use forces us to ask: What will happen if the entire population of the earth consumes resources at the rate of high-wage countries? Neoclassical economists generally ignore this question or give the facile response that there are no limits.

The steady-state economic paradigm suggests a different answer. The regenerative and assimilative capacities of the biosphere cannot support even the current levels of resource consumption, much less the manyfold increase required to generalize the higher stan-



RAISING THE INCOMES in the more populous, less wealthy nations will be difficult. Over the next 40 years, the population will double. To reach the higher level of per capita income, the low- and middle-income countries would have to increase their use of resources by a factor of almost 36 (21 × 2 × 0.85). To avoid augmenting the damage to the environment, they would need to boost resource-use efficiency by the same factor.

dards worldwide. Still less can the ecosystem afford an ever growing population that is striving to consume more per capita. As a species, we already preempt about 40 percent of the land-based primary product of photosynthesis for human purposes. What happens to biodiversity if we double the human population, as we are projected to do over the next 30 to 50 years?

These limits put a brake on the ability of growth to wash away the problems of misallocation and maldistribution. In fact, free trade becomes a recipe for hastening the speed with which competition lowers standards for efficiency, distributive equity and ecological sustainability.

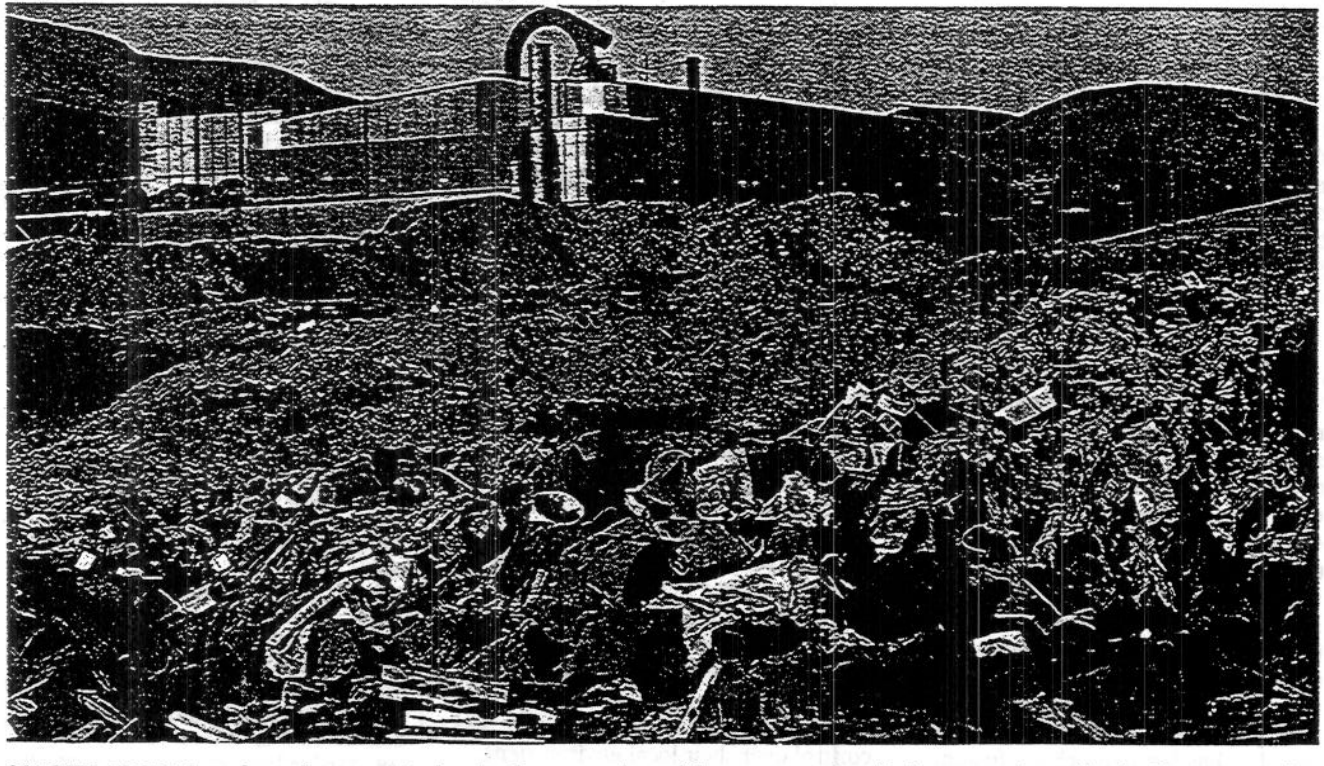
Notwithstanding those enormous problems, the appeal of bigger free trade blocs for corporations is obvious. The broader the free trade area, the less answerable a large and footloose corporation will be to any local or even national community. Spatial separation of the places that suffer the costs and enjoy the benefits becomes more feasible. The corporation will be able to buy labor in the low-wage markets and sell its products in the remaining high-wage, high-income markets. The larger the market, the longer a corporation will be able to avoid the logic of Henry Ford, who realized that he had to pay his workers enough for them to buy his cars. That is why transnational corporations like free trade and why workers and environmentalists do not.

In the view of steady-state economics, the economy is one open subsystem in a finite, nongrowing and materially closed ecosystem. An open system takes matter and energy from the environment as raw materials and returns them as waste. A closed system is one in which matter constantly circulates internally while only energy flows through. Whatever enters a system as input and exits as output is called throughput. Just as an organism survives by consuming nutrients and excreting wastes, so too an economy must to some degree both deplete and pollute the environment. A steady-state economy is one whose throughput remains constant at a level that neither depletes the environment beyond its regenerative capacity nor pollutes it beyond its absorptive capacity.

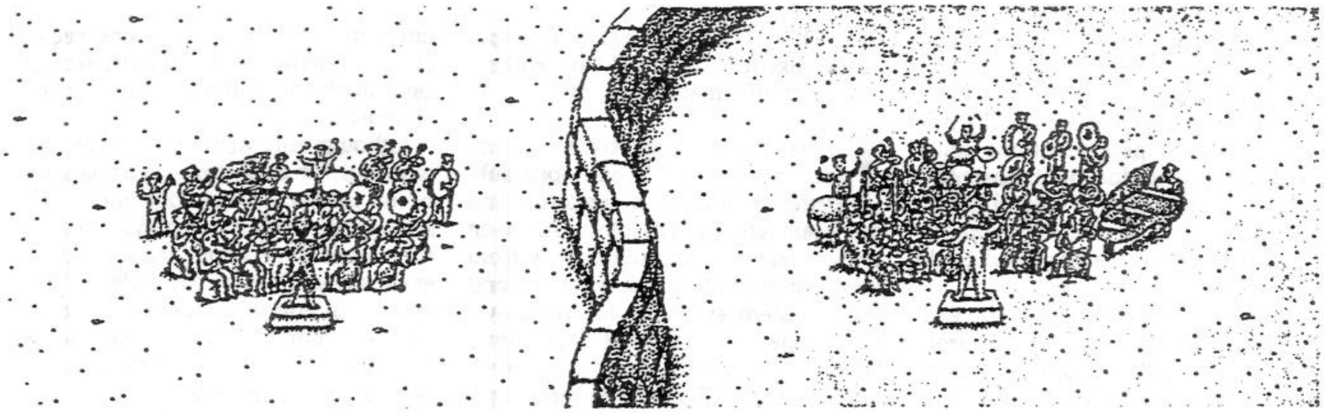
Most neoclassical economic analyses today rest on the assumption that the economy is the total system and nature is the subsystem. The economy is an isolated system involving only a circular flow of exchange value between firms and households. Neither matter nor energy enters or exits this system. The economy's growth is therefore unconstrained. Nature may be finite, but it is seen as just one sector of the economy, for which other sectors can substitute without limiting overall growth.

Although this vision of circular flow is useful for analyzing exchanges between producers and consumers, it is actively misleading for studying scale—the size of the economy relative to the environment. It is as if a biologist's vision of an animal contained a circulatory system but not a digestive tract or lungs. Such a beast would be independent of its environment, and its size would not matter. If it could move, it would be a perpetual motion machine.

Long ago the world was relatively empty of human beings and their belongings (man-made capital) and relatively full of other species and their habitats (natural capital). Years of economic growth have changed that basic pattern. As a result, the limiting factor on future economic growth has changed. If man-made and natural capital were good substitutes for one another, then natural capital could be totally replaced. The two are complementary, however, which means that the short supply of one imposes limits. What good are fishing boats without populations of fish? Or sawmills without forests? Once the number of fish that could be sold at market was primarily limited by the number of boats that could be built and



MAQUILADORAS, or factories near the border between the U.S. and Mexico, have become a troublesome source of pollution for that area. Some U.S. manufacturers have built such factories in Mexico to take advantage of that country's lower labor costs and pollution-control standards. If commerce becomes less regulated, such problems may become more common. Mexican environmentalists closed this plant after showing that it was contaminating its vicinity with lead.



NATIONAL SELF-SUFFICIENCY is a good commonly overlooked by free traders. Just as nations are better off having their own symphony orchestras and other cultural offerings, they should also keep their vital industries local.

manned; now it is limited by the number of fish in the sea.

As long as the scale of the human economy was very small relative to the ecosystem, no apparent sacrifice was involved in increasing it. The scale of the economy is now such that painless growth is no longer reasonable. If we see the economy as a subsystem of a finite, nongrowing ecosystem, then there must be a maximal scale for its throughput of matter and energy. More important, there must also be an optimal scale. Economic growth beyond that optimum would increase the environmental costs faster than it would the production benefits, thereby ushering in an antieconomic phase that impoverished rather than enriched.

One can find disturbing evidence that we have already passed that point and, like Alice in *Through the Looking Glass*, the faster we run the farther behind we fall. Thus, the correlation between gross national product (GNP) and the index of sustainable economic welfare (which is based on personal consumption and adjusted for depletion of natural capital and other factors) has taken a negative turn in the U.S.

Like our planet, the economy may continue forever to develop qualitatively, but it cannot grow indefinitely and must eventually settle into a steady state in its physical dimensions. That condition need not be miserable, however. We economists need to make the elementary distinction between growth (a quantitative increase in size resulting from the accretion or assimilation of materials) and development (the qualitative evolution to a fuller, better or different state). Quantitative and qualitative changes follow different laws. Conflating the two, as we currently do in the GNP, has led to much confusion.

Development without growth is sustainable development. An economy that is steady in scale may still continue to develop a greater capacity to satisfy human wants by increasing the efficiency of its resource use, by improving social institutions and by clarifying its ethical priorities—but not by increasing the resource throughput.

In the light of the growth versus development distinction, let us return to the issue of international trade and consider two questions: What is the likely effect of free trade on growth? What is the likely effect of free trade on development?

Free trade is likely to stimulate the growth of throughput. It allows a country in effect to exceed its domestic regenerative and absorptive limits by "importing" those capacities from other countries. True, a country "exporting" some of its carrying capacity in return for imported products might have increased its throughput even more if it had made those products domestically. Overall, nevertheless, trade does postpone the day when countries must face up to living within their natural regenerative and absorptive capacities. That some countries still have excess carrying capacity is more indicative of a shortfall in their desired domestic growth than of any conscious decision to reserve that capacity for export.

By spatially separating the costs and benefits of environmental exploitation, international trade makes them harder to compare. It thereby increases the tendency for economies to overshoot their optimal scale. Furthermore, it forces countries to face tightening environmental constraints more simultaneously and less sequentially than would otherwise be the case. They have less opportunity to learn from one another's experiences with controlling throughput and less control over their local environment.

The standard arguments for free trade based on comparative advantage also depend on static promotions of efficiency. In other words, free trade in toxic wastes promotes static efficiency by allowing the disposal of wastes wherever it costs less according to today's prices and technologies. A more dynamic efficiency would be served by outlawing the export of toxins. That step would internalize the disposal costs of toxins to their place of origin—to both the firm that generated them and the nation under whose laws the firm operated. This policy creates an incentive to find technically superior ways of dealing with the toxins or of redesigning processes to avoid their production in the first place.

All these allocative, distributional and scale problems stemming from free trade ought to reverse the traditional default position favoring it. Measures to integrate national economies further should now be treated as a bad idea unless proved otherwise in specific cases. As Ronald Findley of Columbia University characterized it, comparative advantage may well be the "deepest and most beautiful result in all of economics." Nevertheless, in a full world of internationally mobile capital, our adherence to it for policy direction is a recipe for national disintegration.

FURTHER READING

FOR THE COMMON GOOD. H. E. Daly and J. B. Cobb, Jr. Beacon Press, 1989.

INTERNATIONAL TRADE AND THE ENVIRONMENT. Edited by Patrick Low. World Bank, 1992.

POPULATION, TECHNOLOGY AND LIFESTYLE: THE TRANSITION TO SUSTAINABILITY. Edited by Robert Goodland et al. Island Press, Washington, D.C., 1992.

MYTHS AND MISCONCEPTIONS OF FREE TRADE. Ravi Batra. Scribner's, 1993.

INTERNATIONAL TRADE AND ENVIRONMENT. Edited by Carl Folke et al. Special issue of *Ecological Economics*, Vol. 9, No. 1; February 1994 (in press).

37
WHARTON REPROGRAPHICS

LIPPINCOTT LIBRARY
DOCUMENT DELIVERY SERVICE

WORLD ECONOMY 14(4), 1991

Managing Trade and Environment: The Demand for Raw Ivory in Japan and Hong Kong

Edward B. Barbier

1. INTRODUCTION

IN the 1980s the population of elephants in Africa halved in eight years from 1.2 million to just over 600,000 (Douglas-Hamilton 1989; and IUCN/WWF/UNEP, 1982).[1] As the elephant's decline has been largely attributed to illegal harvesting (i.e. 'poaching') of elephants for the international ivory trade, in October 1989 signatories to the Convention on International Trade of Endangered Species (CITES) voted for an indefinite ban in all trade in elephant products, with the exception of trophy imports from safari hunting. An amendment also made it illegal for countries to trade from their existing stockpiles of raw ivory, which particularly affects countries such as Burundi and Hong Kong that have accumulated large stockpiles.[2]

Although the majority of the African nations and ivory importing countries have supported the CITES ban, the policy debate is far from over. First, some key exporting countries (e.g. Zimbabwe, Botswana, South Africa, Malawi and Zambia) and importing countries (e.g. Hong Kong and China) have taken treaty reservations against the ban and continued trading in ivory. In addition, critics

---

EDWARD B. BARBIER is from the London Environmental Economics Centre, IIED and University College, London. He is grateful to Joanne C. Burgess, Timothy M. Swanson, David W. Pearce, Ron P. Smith, John Whalley and two anonymous referees. Any remaining errors are the author's own.

[1] However, in a few African countries — South Africa, Botswana, Zimbabwe and Namibia — elephant populations have stabilised or risen.

[2] Although indefinite, the CITES ban is not complete. In the future, individual African countries that can demonstrate having healthy elephant populations, workable elephant conservation and management programmes and effective ivory trade controls may be 'certified' by a 'technical committee' (yet to be established) to re-engage in a limited trade.

have argued that a complete ban is not an appropriate way of saving the elephan (Parker, 1989; and Simmons and Kreuter, 1989). An alternative international regulatory system, or 'Ivory Exchange', involving an 'enforceable' producers' cartel, limited ivory trade, and possible import taxes, has been proposed (Barbie et al., 1990; and Swanson, 1989). CITES itself will be reviewing its own position periodically, beginning in 1992, and it is anticipated that all policy options will be reviewed and considered.

From a trade policy perspective, the ivory ban is a good example of a commodity being subjected to severe trade restrictions on environmental grounds, i.e. conservation of elephant populations. The ban therefore provides an interesting case study in the use of trade policy instruments to achieve an environmental objective — a practice that may become increasingly more prevalent for a whole range of commodities. Two important issues emerge: first, are trade policy interventions the appropriate instruments for achieving an environmental objective, and second, if trade interventions are justified, what is the optimal instrument for achieving this objective?

A recent study of the economics of the ivory trade has argued tha international regulation is required, but an indefinite trade ban is not the appropriate response (Barbier *et al.*, 1990). A key factor identified by the study in determining the optimal policy is the role of demand by importing countries. The remainder of this paper examines the trends in total and net demand for unworked (raw) ivory of Japan and Hong Kong – the two major ivory importers before the ban – over the periods 1950–87 and 1962–87, respectively. Demand analysis was conducted through two econometric models. The first estimates income and price elasticities of demand at a national level. The second also includes an 'asset demand' motivation for holding ivory, which is asserted by some authors as being significant (Douglas-Hamilton, 1987). The results of the analysis suggest that control of end-use demand through a combination of regulatory measures and taxes may be the most appropriate policy, although much more analysis is required of the inter-relationship between trade in worked and unworked ivory and of the welfare implications of the global trade.

2. THE ROLE OF THE IVORY TRADE

Total exports of unworked ivory leaving Africa have risen from between 200 to 400 tonnes per annum (t/yr.) in the 1950s to around 1000t in 1980. This represents a greater than 400 per cent increase in the trade over 40 years, or a 10 per cent increase in each year of the past four decades. Up to 1986, records indicate that between 700 and 1,000 tonnes of raw ivory were traded internationally each year (see Table 1).[3] The sum total of ivory exported between 1979 and 1988 amounted to between 7,625 and 7,955t, accounting for approximately 700,000 to 800,000 elephants. Over the same 1979 to 1988 period, the price of raw ivory increased dramatically from around $60 to $150 per kilogramme. In early 1989 – before moves to impose an international ban — the price rose towards $300 per kilogramme (Barbier *et al.*, 1990). Thus the ivory trade has been a highly lucrative activity.

TABLE 1
Raw Ivory Exports by Major African Exporters, 1979-88
(tonnes)

	1979	*1980*	*1981*	*1982*	*1983*	*1984*	*1985*	*1986**	*1987**	*1988**
Botswana	13	5	9	5	6	4	17	0.4	0.4	0.1
Burundi	139	126	61	46	124	184	215	138	50	9
CAR	181	183	107	202	204	122	117	19	2	0.3
Chad	27	4	13	28	34	4	0.03	0	2	0
Congo	93	175	237	102	53	95	72	17	86	19
Gabon	4	1	1	0.2	0.9	0.2	0.4	0.5	4	14
Kenya	46	31	6	12	4	13	19	2	0.1	0
S. Africa	42	35	33	28	46	36	50	41	18	8
Sudan	124	206	269	278	337	68	22	78	70	0
Tanzania	33	45	27	18	15	45	114	302	56	43
Uganda	25	19	45	12	13	100	206	3	0	0.2
Zaire	160	84	45	78	157	90	22	23	12	11
Zambia	16	22	28	34	18	2	14	11	4	2
Zimbabwe	3	2	1	14	14	21	23	8	8	7
All Africa Min. Total	944	952	905	890	1032	798	912	719	331	142
All Africa Max. Total	950	1164	912	900	1040	798	912	805	331	142

Note: These are the 'best estimates' of African exports, disaggregated by country of export, using import and export data from customs sources, and CITES data. Both *legal* and *illegal* exports are likely to be included. For starred years (*), because of the adoption of the management quota system by CITES, there may have been an incentive to evade both customs and CITES systems in some countries.

Source: E.B. Barbier, J.C. Burgess, T.M. Swanson and D.W. Pearce (1990). *Elephants, Economics and Ivory* (Earthscan Publications, London), Table 1.2.

The annual value of ivory exports from Africa in the 1980s is estimated to be around $50–60 million (Barbier *et al.*, 1990). As this includes c.i.f. costs and marketing margins in international trade, an estimate of $35–45 million is probably a more realistic measure of the value of African ivory exports. This amount is a tiny fraction of all African export earnings, although important for a few individual countries. Recent studies suggest that other values of the elephant, such as its importance to tourism earnings, may be considerably more significant.[4]

---

[3] The decline in tonnage since the mid 1980s shown in Table 1 looks encouraging, but the statistics hide the real impact of the ivory trade on the elephant population. By 1987 most of the mature bull elephants had been shot, leaving cows and calves with much smaller tusks to support the demand for ivory. On average, one tonne of traded ivory represented about 113 dead elephants in 1987 as opposed to 54 in 1979. As a consequence, almost the same number of elephants were harvested in 1987 as in 1979 to support the total demand for ivory, and from a much reduced and more fragile population (Barbier et al., 1990). Note also that many of the exporters listed in Table 1 do not have significant elephant populations of their own. Several countries have acted as entrepots for their neighbours, notably Sudan, Burundi, the Central African Republic (CAR), and Congo, each of which exported more than 900t of ivory between 1979 and 1988. Much of this entrepot trade involved illegally harvested ivory, which gives some indication of the magnitude of the activity.

The flow of ivory in international trade has historically been toward the hard currency consumers – the USA, Europe and, more recently, Japan. These three consumers acquired about 75 per cent of all ivory worked in Hong Kong and Japan in the 1980s. The remainder of the ivory in international trade has gone to the carving industries within various developing countries – for example, India and China – and then been re-exported or sold to tourists for hard currency. The same is true for much of African worked ivory. It is very likely therefore, that nearly all the ivory in international trade has ultimately found its way to one of the three hard currency consumers: the United States, Europe and Japan.

TABLE 2
Net Imports of Raw Ivory by Major Consumers
(tonnes)

	1979	*1980*	*1981*	*1982*	*1983*	*1984*	*1985*	*1986**	*1987**	*1988**
USA	6	23	11	7	20	55	24	17	21	9
FR Germany	74	181	32	35	43	–7	16	7	2	1
UK	–5	–26	0	–3	2	3	28	–1	7	3
Hong Kong	366	376	427	318	428	267	109	129	150	133
India	17	19	19	24	23	30	21	8	6	4
France	89	22	7	4	11	21	5	5	4	–2
China	7	10	10	54	20	7	7	19	39	50
Japan	270	240	256	205	174	179	206	29	103	75
Thailand	1	1	2	4	–5	–12	–2	1	0	–3
Belgium	16	–90	–248	–123	–105	–116	0	0	–10	12
Singapore	–7	–4	3	7	0	120	60	324	–148	–129
Macau		0	0	5	16	38	82	57	8	11
Taiwan	11	18	17	18	28	34	21	18	80	5
Total Net Imports, All Countries (Min. estimate)	979	967	895	891	1018	710	749	600	370	153

Note: These 'best estimates' of final demand for raw ivory by the main consuming countries are based on customs trade statistics and CITES documentation.
Minus figures indicate that countries were *net exporters* in that year, most likely through de-stocking.
For starred years (*), because of the adoption of the management quota system by CITES, there may have been an incentive to evade both customs and CITES systems in some countries.

Source: E.B. Barbier, J.C. Burgess, T.M. Swanson and D.W. Pearce (1990). *Elephants, Economics and Ivory* (Earthscan Publications, London), Table 1.3.

---

[4] For example, Brown and Henry (1989) estimate through a travel-cost study the value of viewing elephants in Kenya to be around US$25 million per year — as much as ten times the value of its (poached) ivory exports.

The primary cause of the growth in the ivory trade may be traced to the ivory's character as a precious material, used in sculpture and jewellery. Ivory has been a culturally demanded substance in Asia for several centuries. In Japan the ivory is especially prized for making *hankos*, personal seals traditionally used by some Japanese in place of a signature. Tables 2 and 3 confirm that the major demand for raw and worked ivory since 1979 has been in East Asia, principally in Japan and Hong Kong.

TABLE 3
Net Imports of Raw and Worked Ivory by Major Consumers

Net Imports of Raw Ivory by Major Consumers
(% of Total)

	1979	*1980*	*1981*	*1982*	*1983*	*1984*	*1985*	*1986*	*1987*	*1988*
Hong Kong	37	39	48	36	40	38	15	22	40	86
Japan	28	25	32	29	20	30	27	5	28	49
EEC	18	12	5	5	6	4	6	2	1	9
USA	1	3	1	1	2	8	3	3	6	6
Taiwan	2	2	2	2	3	3	3	3	22	3
Macau	0	0	0	0	2	4	11	10	2	7
China	1	1	1	7	2	0	1	3	10	32
India	2	2	2	3	2	3	3	1	2	3
Singapore	(–1)	0	0	1	0	17	8	54	(–40)	(–84)
Thailand	0	0	0	0	0	(–2)	0	0	0	(–2)

Source: Table 2.

Net Imports of Worked Ivory by Major Consumers
(% of Total, 1979-84 average)

Japan	38
EEC	18
USA	16
Other Asian (Singapore, Taiwan, S. Korea and Macau)	3

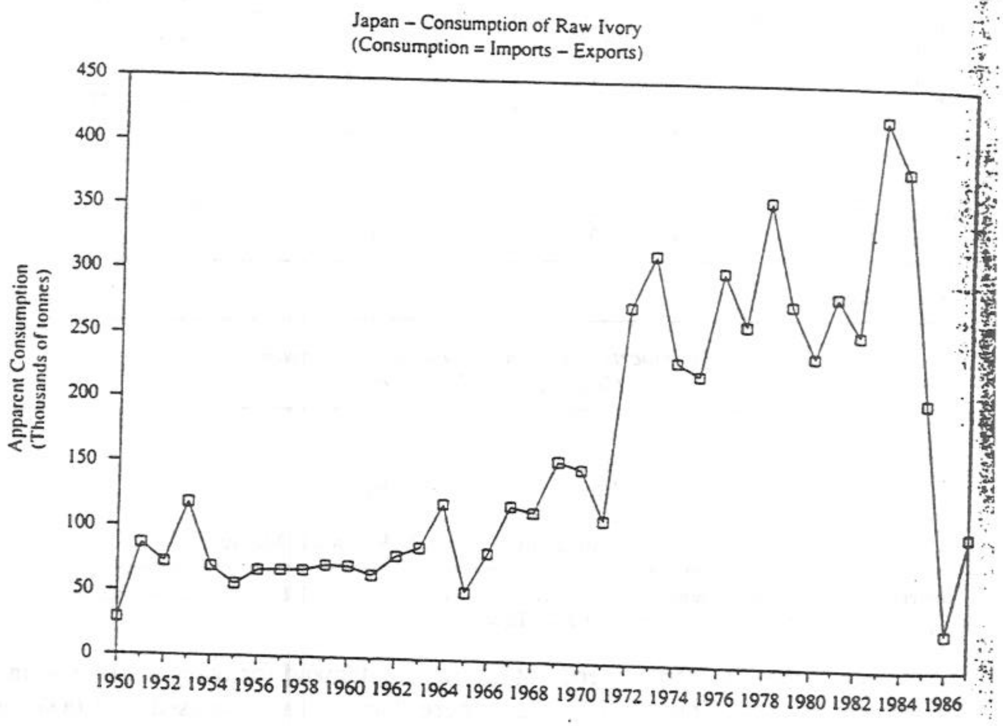
Source: E.B. Barbier, J.C. Burgess, T.M. Swanson and D.W. Pearce (1990). *Elephants, Economics and Ivory* (Earthscan Publications, London), Table 1.4.

However, there is also evidence of increasing demand for ivory in other Asian countries – such as Korea and Taiwan, where demand has increased by 1,000 per cent in the past decade – and that demand pressure will remain high so long as rapid economic growth continues in this region. Other rapidly developing countries, such as Singapore, Dubai and Macau, have also been playing an increasingly important role in the movement of ivory stocks and in their processing and carving. This fluidity represents one of the characteristics of both the legal and illegal trade: whenever one channel through which ivory is passing has been blocked, another has opened up almost immediately. The entrance of new countries into the ivory trade and its fluidity are important issues to be tackled by any regime of international regulation. Demand analysis of the import market in Japan and Hong Kong provides important policy implications for designing such a regime.

3. DEMAND FOR IVORY IN JAPAN AND HONG KONG

Figure 1 indicates the long-term trends in the demand for raw ivory by Japan. Japan's end-use demand, or final consumption, of raw ivory remained fairly stable in the 1950s and early 60s, but then increased dramatically from the late 1960s to the early 1980s.[5] For example, as its peak in 1983 raw ivory consumption was 15 times its level in 1950.

FIGURE 1



In the early 1980s, the relative importance of Hong Kong as the centre of the ivory trade was overtaken by Japan. Even though by 1980 Japan had accepted the control of imported tusks agreed under CITES, up to 1985 Hong Kong's legislation was more effective than Japan's in implementing CITES rules for raw ivory. As a result, illegitimate suppliers of ivory from Africa reduced their exports to Hong Kong, which subsequently became more reliant upon Japan as a source of supply (Caldwell, 1984 and 1987).

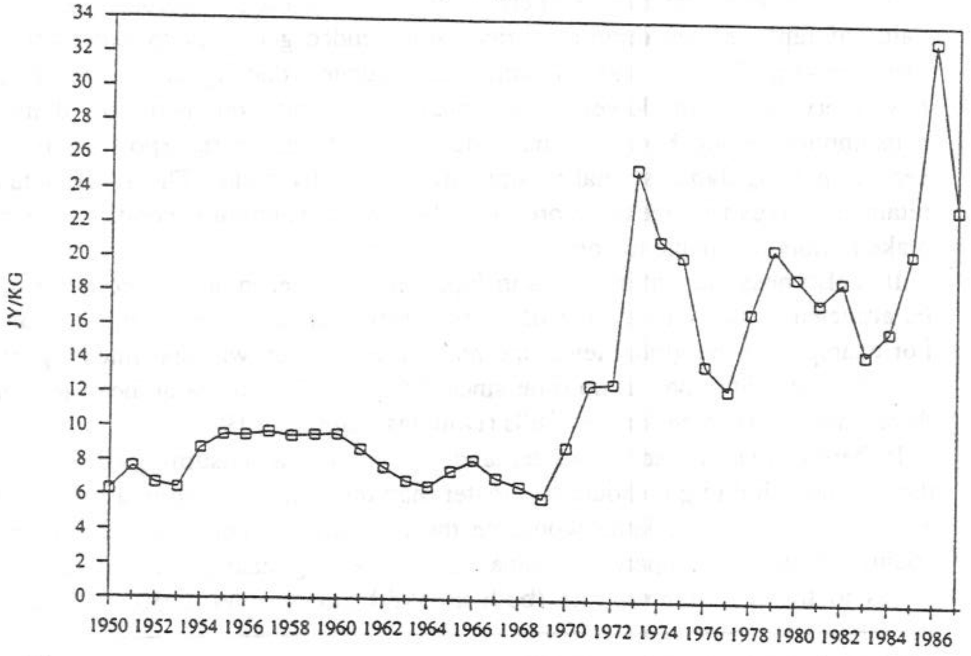
---

[5] Consumption of raw ivory by Japan is equivalent to its net imports, i.e. gross imports less re-exports in the same year. Unfortunately, custom data on re-exports by Japan were only available from 1970 onwards. As noted below, however, only in the 1980s have Japan's re-exports of raw ivory increased significantly (mainly to Hong Kong). Also note that for the 1984-87 period ivory waste and powder imports have been ignored.

However, in 1985 the Japanese Government and ivory importers introduced regulations that conformed to CITES recommendations (Milliken, 1989). The new regulation began affecting net raw ivory imports in the second half of 1985, and in 1986 net imports fell to their lowest level in 20 years (Figure 1). Although there has been a slight recovery in 1987 and 1988, Japan's consumption of raw ivory before the ban remained at around 100 tonnes or even less (Table 2).

FIGURE 2

Japan — Real Raw Ivory Prices
(1980 P)



Given that Japan is a major end-use consumer of ivory, the main effect of the 1985 Japanese regulations was to create a block at the importing end of the illegal trade route. This may have given rise to the diversion of illegal ivory to entrepot and importing consumers that have a less strict control on exports, as evidenced by the subsequent large-scale stockpiling of illegal ivory in Macau, Singapore, and Dubai (Milliken, 1989).[6] Another effect of the 1985 Japanese regulations was to make the illegal traffic from Africa more 'transparent'; i.e., it became easier to trace the tremendous volumes of ivory moving outside CITES channels.[7]

---

[6] In recent years there has been a flourishing illegal trade: via Burundi to the United Arab Emirates and probably on to India; the emergence of Singapore and Dubai as important entrepots for ivory trading without CITES permits; and the emergence of Taiwan and the resurgence of China, and possibly of India, as important destinations of the trade (Parker, 1989).

The price trends over 1950-87 for the ivory imported into Japan reflected the growing scarcity of good quality raw ivory pieces in the world trade. Real raw ivory prices – the per unit value of net imports of raw ivory deflated by the consumer price index – were significantly high, but more volatile, in the 1970s and 80s compared to the 1950s and 60s (Figure 2). Increasing demand for imported ivory in Japan clearly has had a long-term influence on price trends. Over 1983-86, Japanese real raw ivory prices rose sharply but then fell back in 1987. The implications are that the imposition of trade controls over 1985-86 may have cut off the supply of illegal ivory, driving up the price of legal ivory and as a consequence the price of imported ivory.

Moreover, over the 1970-87 period for which data are available, the per unit value of Japanese net imports of raw ivory tended generally to exceed that of gross imports (Barbier, 1989b). This would suggest that Japanese re-exports of raw ivory were of lower value than raw ivory retained for domestic consumption. Hong Kong, the main destination for these re-exports, may have been importing Japan's smaller and inferior quality tusks. The ivory actually retained in Japan might therefore be a 'luxury' consumption good, which may make it more amenable to control by taxation.

It is also possible that investors in Japan and in other importing countries may be attracted to ivory as a store of wealth, especially in times of high inflation. For example, at the global level, the real value of ivory was six times higher in 1978 than in 1960, and at no time since 1973 has the real value been less than three times its average for the 1960s (Douglas-Hamilton, 1987).

If there is a significant 'asset demand' for ivory in a consuming country, then the returns to holding it should be greater than those obtained from an alternative asset. A proxy for the latter would be the real rate of interest in the importing country. Figure 3 compares for Japan the percentage change in real raw ivory prices to the real interest rate (both at 1980 prices) over 1950-87. The real interest rate fell steadily since 1950, whereas the percentage change in real raw ivory prices fluctuated significantly, particularly since 1970. If an 'asset demand' for raw ivory exists, it may be cyclical; i.e., only over certain years may the capital gains from holding ivory exceed those earned from other assets.

---

[7] Before the new regulations, the Japanese import system unwittingly assisted the illegal trade: country of origin certificates were accepted in lieu of the required CITES documents, so that the 'act of smuggling' merely involved moving ivory into Japan openly with documentation which was not sanctioned under less strict import controls. This meant that large-scale 'smuggling' was unnecessary. After 1985 it was possible, although somewhat doubtful, that smuggling increased. On the other hand, traded ivory with proper permits began selling for several times the price of ivory without them (Caldwell, 1987), which may have encouraged producer countries in Africa to maintain a regular supply of legally traded ivory to importing countries, like Japan, that began following CITES rules. A modest growth in the trade between Japan and Hong Kong in cut and simply worked ivory also emerged after 1985, although this apparently halted in 1988 when imports dropped by about 15 per cent. This trade was most likely an attempt to evade the regulations imposed on raw ivory by diverting trade into worked ivory, thus exploiting a loophole in CITES rules.

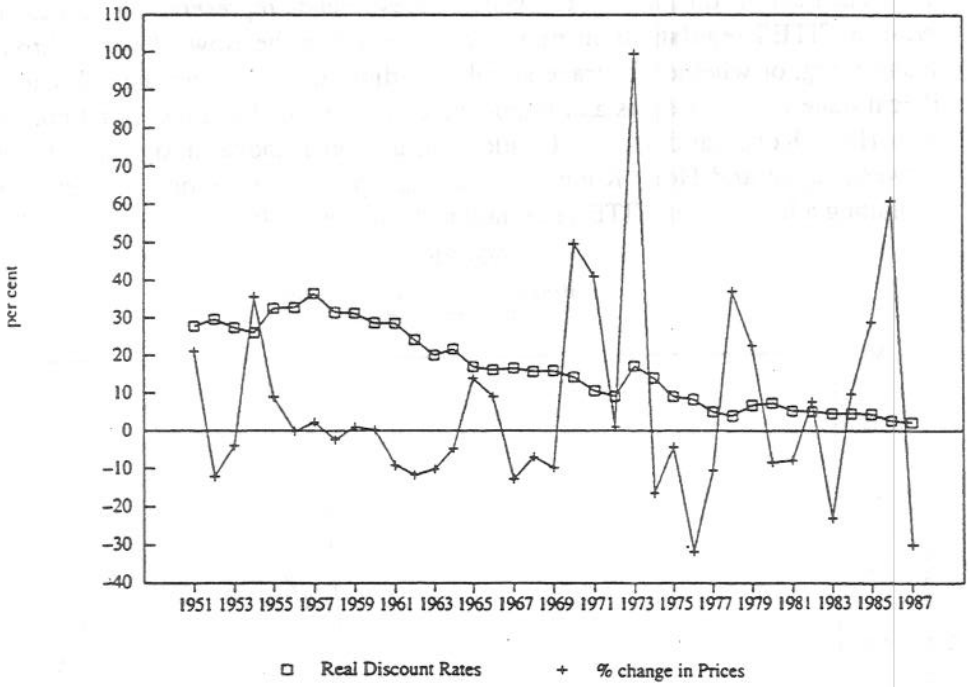


FIGURE 3

Japan: Change in Real Raw Ivory Prices
(1980 P)

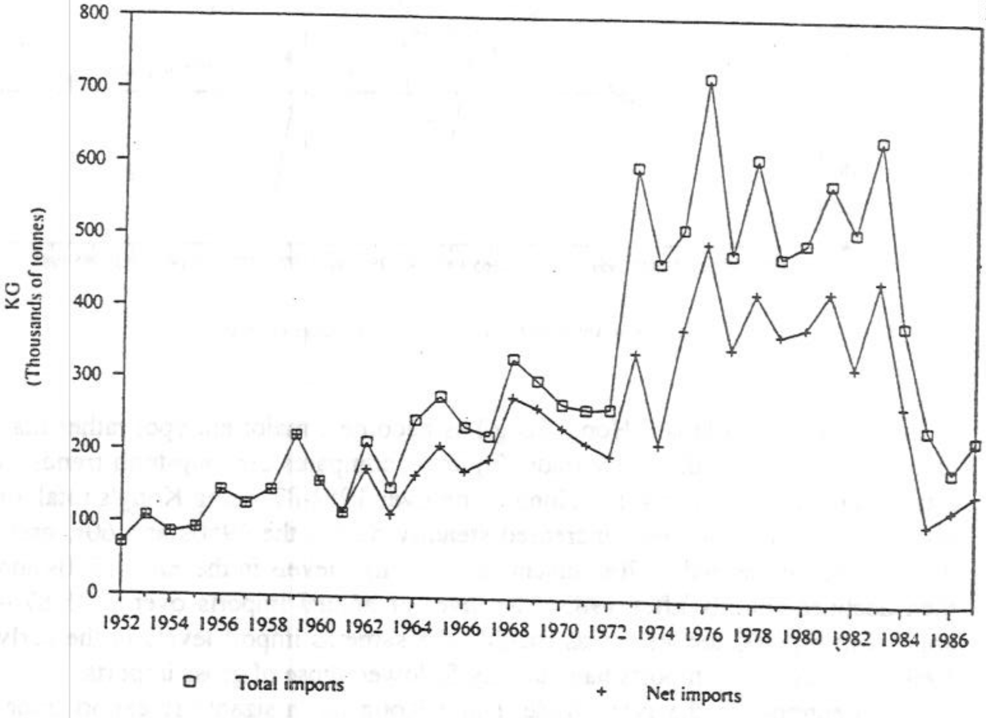
In comparison to Japan, Hong Kong has become a major entrepot rather than end-use consumer in the ivory trade. Figure 4 compares the long-term trends in total and net ivory imports into Hong Kong over 1952-87. Hong Kong's total, or gross, imports of raw ivory increased steadily during the 1950s and 60s, grew dramatically in the mid-1970s, sustained these high levels in the early 1980s and then declined sharply after 1983. The level of annual imports over 1985-87 – approximately 200 tonnes – was roughly the same as import levels in the early 1960s. Trends in net imports have largely followed those of gross imports.

As an entrepot in the ivory trade, Hong Kong has a sizable re-export trade. The share of Hong Kong's re-exports of raw ivory in total imports by weight was roughly 20-25 per cent in the early 1960s, dipped to under 20 per cent in the late 1960s and early 1970s, and, with the exception of unusually strong performances in 1973-74 and 1983, remained at roughly 25-30 per cent ever since (Barbier, 1989a).

In the 1980s the pattern of Hong Kong's trade in raw ivory underwent some important changes. In the early 1980s, the relative importance of Hong Kong as the centre of the ivory trade was overtaken by Japan. A major reason cited is that up to 1985 Hong Kong was more effective than Japan in implementing legislation agreed under CITES (Caldwell, 1984 and 1987). Two results have emerged. First, after 1984 total imports of raw ivory into Hong Kong fell off dramatically (Figure 4). Unfortunately, the imposition of the ivory trade ban in 1989 has made it difficult to tell whether these trends represented a success for pre-ban CITES regulations in permanently reducing the flow of ivory through Hong Kong, or whether the trade simply re-adjusted – i.e., diverting ivory to the illegal trade or to entrepots and importing countries that have less strict controls than Hong Kong (and Japan). In addition, as noted above, a significant trade between Japan and Hong Kong in cut and simply worked ivory also emerged, exploiting a loophole in CITES regulations (Caldwell, 1987).

FIGURE 4

Hong Kong – Imports of Raw Ivory
Total and Net Imports

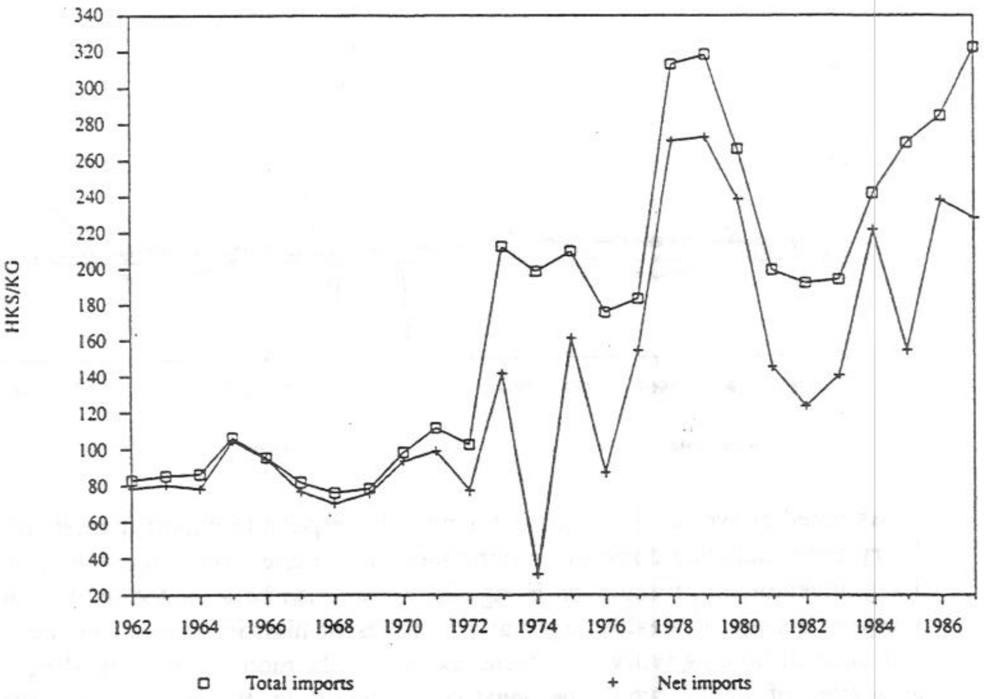


Second, in the late 1980s legitimate suppliers of ivory from Africa reduced their exports to Hong Kong, which subsequently became more reliant on Japan and other non-African suppliers. Thus the long-term trend of Hong Kong importing less ivory direct from African suppliers and increasingly from the rest of the world – in particular Japan – seems to have accelerated in the 1980s (Barbier, 1989a).

In Hong Kong, real raw ivory prices – the per unit value of total imports of raw ivory deflated by the consumer price index – rose sharply in the early 1970s and again in the late 70s, but fell in the early 1980s before returning recently to maximum levels. Prices for net imports have generally followed similar trends (Figure 5). With the exception of the 1983-87 period, it is not apparent that raw ivory imports are immediately responsive to changes in real prices, although some evidence for a lagged impact seems present (Figures 4 and 5). The latter would be consistent with the stockpiling of ivory.

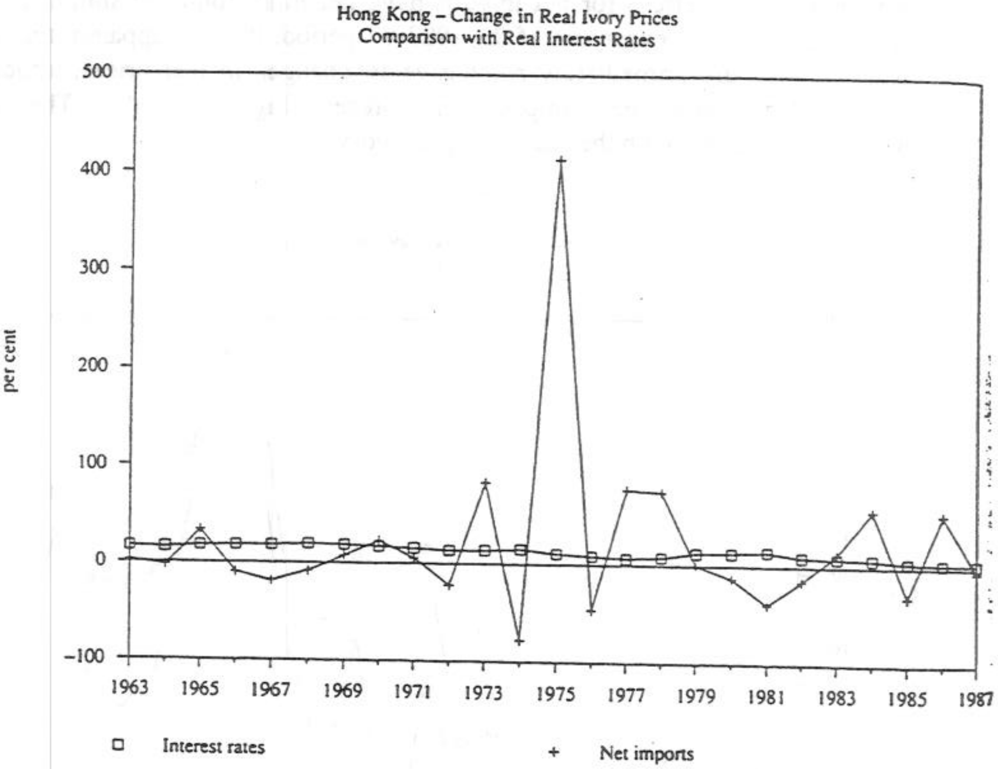
FIGURE 5

Hong Kong – Real Raw Ivory Prices
(1979/80 P)



In Japan, the per unit value of net imports of raw ivory has tended generally to exceed that of gross imports. This is consistent with a final consuming country that re-exports lower value raw ivory, retaining the higher valued ivory for domestic consumption. In contrast, in Hong Kong, over the period 1962-87 for which re-export data are available, the per unit value of total imports of raw ivory tended to exceed that of net imports (Barbier, 1989a). Thus, with a substantial re-export market in raw ivory and an export-oriented carving industry, Hong Kong appears to have been exporting higher valued raw ivory and retaining the rest as raw material for its carvers. This suggests that the responsiveness of both gross and net imports to price changes might be different for an entrepot as opposed to an end-use consumer of ivory.

FIGURE 6


Hong Kong – Change in Real Ivory Prices
Comparison with Real Interest Rates

As noted above, in Hong Kong the need to respond to export markets of raw ivory and to supply a domestic carving industry requires some stockpiling of raw ivory inventories. Investors in Hong Kong may also be attracted to stockpiling ivory as a store of wealth, especially in times of high inflation. Thus both the real costs of holding ivory inventories as well as the motivation for holding ivory as a store of wealth would be negatively affected by the level of real interest rates in the economy. Figure 6 compares for Hong Kong the percentage change in real raw ivory prices for net imports to the real interest rate (both at 1979/80 prices). The real interest rate has fallen steadily since 1963, whereas the changes in prices have been generally more cyclical. This would suggest that if there is any impact of real interest rates on the stockpiling of ivory, it may be cyclical.

Moreover, if stockpiling, including 'asset demand,' is a significant component of the total demand for ivory in Hong Kong, this would explain the considerable fluctuations in imports as indicated in Figure 4.

4. DEMAND ANALYSIS: JAPAN AND HONG KONG

An important policy option that needs to be explored is whether the key to regulating the world trade in raw ivory lies in controlling the demand for ivory by end-use countries, such as Japan, and/or by entrepot countries, such as Hong Kong. In the case of Hong Kong, this demand consists not only of end-use domestic consumption of ivory – which may actually be quite small – but also of supplying its re-export trade and its export-oriented carving industry, as well as any stockpiling that is required to service both. The demand for ivory in Hong Kong and other entrepot countries therefore differs significantly from that in mainly end-use consuming countries, such as Japan. Analysis of the demand for ivory by end-use and entrepot countries should reveal these differences. Such an analysis should also indicate whether it is feasible for entrepot and consumer countries such as Hong Kong and Japan to employ a tax, either on gross or net imports, to reduce their demand for ivory.

TABLE 4
Summary of Elasticities of Demand for Raw Ivory — Japan

	Model 1 *No interest rate variable, net imports*	*Model 2* *With interest rate variable, net imports*	
		With constant	*Without constant*
Price Elasticity			
i. current year	−0.52	−0.70	−0.72
Income Elasticity			
i. current year	+1.00	+1.15	+1.53
Interest Elasticity			
i. current year	n/a	+0.34	+0.37
ii. lagged 1 year	n/a	−0.40	−0.26

Source: E.B. Barbier (1989b). 'The Demand for Unworked Ivory: A Case Study of Japan', Economics Working Paper ITRG/EG 89-01, London. Published in E.B. Barbier and J.C. Burgess (1989). *The Demand for African Elephant Ivory*, LEEC Paper 89-05 (London Environmental Economics Centre, London).

Estimates of the elasticities of demand for ivory in Japan were calculated, based on a Cobb-Douglas model relating Japanese consumption of raw ivory to real raw ivory prices and real gross national product. An additional model was also constructed that included an interest rate variable to analyze the significance of an 'asset demand' motivation for holding ivory. Several versions of the two models were tested. The results are summarised in Table 4, with further details provided in the Appendix.[8] As the effect of interest rates on demand appeared to have been significant, model 2 produced the best estimation of elasticities, especially the 'without constant' version.

The estimates of the elasticities of Japanese end-use demand for raw ivory therefore reveal a price elasticity of about one half and an income elasticity of approximately unity. When an 'asset demand' motivation is included, price elasticity increases to over 0.7 and income elasticity rises to at least 1.15, and in the best model, to over 1.53. Intuitively this seems reasonable. If ivory is being held as a store of wealth, one would expect it to be more income-elastic (i.e., more of a luxury good), and one would also expect demand to become more sensitive to changes in real prices, as these now determine the returns on ivory as an asset.

Although the above models do seem to present reasonable evidence that such an 'asset demand' motivation for holding ivory in Japan does exist, further work is required before one can conclusively reject the hypothesis that no such motivation is at work. Equally important, if such a relationship does exist, more analysis is required before the actual mechanics and magnitude of the 'asset demand' for ivory can be satisfactorily revealed.

As for Japan, estimates of the elasticities of demand for ivory in Hong Kong were calculated, based on a Cobb-Douglas model related Hong Kong's imports of raw ivory to real raw ivory prices, real gross national product and the exchange rate (HK$/US$). An additional model was also constructed that included an interest rate variable to analyze the significance of an 'asset demand' motivation for holding ivory. Several version of the two models were tested, for both *total* and *net* raw ivory imports. The results are summarized in Table 5, with further details provided in the Appendix.[9]

The model estimates indicate the complexity of the relationships determining the demand for raw ivory in Hong Kong. First, there are clearly important differences between total and net ivory demand. Second, the cyclical nature of the demand is reflected in the presence of significant lagged effects. Finally, inclusion of an interest rate variable in the estimations affected other variables dramatically.

---

[8] Full details on the models and the estimation results can be found in Barbier (1989b) and Barbier and Burgess (1989). In the analysis data on Japanese imports of raw ivory were based on Japanese Customs import statistics and not from Japanese CITES data. This has been the procedure followed by past qualitative analysis of the Japanese trade in unworked ivory (Caldwell, 1984 and 1987; Parker, 1979 and 1988; and WTMU, 1983).

[9] Full details on the models and the estimation results can be found in Barbier (1989a) and Barbier and Burgess (1989). As in the case of the Japan analysis, data on Hong Kong imports of raw ivory were based on Hong Kong Customs import statistics and not from Hong Kong CITES data.

TABLE 5
Summary of Elasticities of Demand for Raw Ivory — Hong Kong

	Model 1 *No interest rate variable*		*Model 2* *With interest rate variable*	
	Net Imports	*Gross Imports*	*Net Imports*	*Gross Imports*
Price Elasticity				
i. current price.	n/a	+0.42	+0.48	+0.41
ii. lagged 1 year	–0.51	–0.55	–0.20	–0.41
iii. lagged 2 years	+0.35	n/a	n/a	n/a
Income Elasticity				
i. current year	+4.43	+1.08	+0.40	+1.54
ii. lagged 1 year	–3.52	–0.81	n/a	–1.20
iii. lagged 2 years	–0.86	n/a	n/a	n/a
Exchange Rate Elasticity				
i. current year	+4.7	+1.12	–0.67	+2.10
ii. lagged 1 year	–5.7	–1.41	–1.07	–4.22
iii. lagged 2 years	n/a	–2.08	n/a	n/a
Interest Elasticity				
i. current year	n/a	n/a	+0.49	–0.47
ii. lagged 1 year	n/a	n/a	–0.55	+0.72
iii. lagged 2 years	n/a	n/a	+0.83	n/a

Source: E.B. Barbier (1989a). 'The Demand for Unworked Ivory: A Case Study of Hong Kong', Economics Working Paper ITRG/EG 89-09, London. Published in E.B. Barbier and J.C. Burgess (1989). *The Demand for African Elephant Ivory*, LEEC Paper 89-05 (London Environmental Economics Centre, London).

For both *net* and *total* import demand, exclusion of interest rates from the estimation (model 1) led to extremely high measures of income and exchange-rate elasticities of demand. In model 2, when interest rates were included, not only did the explanatory power of the regressions improve, but the over-estimation biases of the exchange rate and income variables were also reduced. Thus the elasticity estimates of model 2 are preferred. The large fluctuations in the magnitudes and even the signs of the estimates suggest caution in interpreting results.

The income elasticities of model 2 indicate that total imports of raw ivory are more of a 'luxury' good than net imports. As noted above, Hong Kong re-exports its higher valued raw ivory while retaining lower valued ivory to service its carving industry. Thus like most raw material commodities, the raw ivory retained in Hong Kong displays a low income elasticity. In contrast, a large component of Hong Kong's gross ivory imports includes high-valued raw ivory that is largely exported to final consuming countries such as Japan. Thus Hong Kong's total imports are much more income elastic, reflecting its re-export trade in 'luxury' raw ivory.

Similarly, if net exports are used primarily as a raw material or stockpiled for future re-export, then a negative exchange rate elasticity would be expected (see Table 5). Higher exchange rates would raise significantly the costs to the carving industry of using this input, causing net imports of raw ivory to fall. In contrast, if a large amount of gross imports are immediately re-exported, then total imports would expect to increase with a rise in the current exchange rate, which benefits exports.

The elasticity estimates for price and interest rates are more cyclical and difficult to interpret. Given that Hong Kong may be importing large volumes of raw ivory for immediate export, higher real import prices for raw ivory may actually be encouraging total imports to increase to supply the re-export trade. Higher prices in the previous year may discourage importing in the current year as it is anticipated that real prices will be lower. Similarly, a rising net import price for *raw* ivory may actually be a reflection of stronger demand, and higher prices, for *worked* ivory carved in Hong Kong. Thus the domestic ivory carving industry is expanding and importing more raw ivory inputs causing raw ivory import prices to rise. Higher prices today may eventually reduce future net imports, but this lagged effect appears quite small – an elasticity of 0.2.

The changes in the sign of the lagged interest rate variables also indicate that net imports of raw ivory into Hong Kong are responding to the cyclical pattern of the change in real interest rates over time. Thus a higher level of real interest rates in the previous year will reduce net imports of raw ivory today, either because investors will have been attracted to other income-earning assets and thus in the future invest less in ivory as a store of wealth, or because higher real interest rates raise the cost of financing the next year's stockpiling of ivory. On the other hand, high current levels of interest rates today may mean that future rates are expected to be lower and thus ivory becomes more attractive as an alternative asset and stockpiling easier to finance. In contrast, for *total* ivory imports, higher real interest rates would discourage any 'asset demand' motivation for importing additional exportable ivory and holding on to it as a store of wealth. If past real interest rates were rising, however, the expectation would be that future real rates would be lower, which would be an incentive to import.

In sum, the above analysis of net and total imports reaffirms that Hong Kong is not only a key entrepot country in the world trade for raw ivory, but it also retains a substantial amount of raw ivory to supply its export-oriented worked ivory industry. The economic motivation for importing ivory is therefore extremely complex, as reflected in the cyclical and lagged impact of many of the variables.[10]

5. POLICY IMPLICATIONS OF RESULTS

The results of the above models have important policy implications. First, as the demand for raw ivory in Japan seems to be significantly sensitive to real prices as well as income-elastic, there appears to be some scope for the use of import taxes as a means of discouraging domestic consumption. For example, assuming 0.7 price elasticity, a domestic tax on consumption that raised real ivory prices by 10 per cent would reduce the 1987 level of consumption by 7.2 tonnes. Taxes that *doubled* the real price of raw ivory would reduce Japanese consumption by about 70 per cent, which in 1987 would have meant a total net import level of 7,228 kg – about the level of gross imports in Japan in the early 1960s.

As raw ivory demand in Japan appears to be fairly income elastic, it is essential to take measures to reduce consumption as real GNP in Japan continues to expand at around four per cent per annum. This could increase ivory consumption by 6.1 per cent each year – or an additional 6.3 tonnes based on 1987 levels. In addition, the high income elasticity of demand for ivory would suggest that any tax should be progressive. Thus one would anticipate that the higher income groups in Japan would bear proportionately more of the burden of taxation, particularly if there is a strong asset demand motivation for holding ivory.

Imposing high taxes on raw ivory consumption in Japan would not be effective unless accompanied by tight CITES controls and other import legislation that has a bearing on ivory imports. In other words, taxes alone would not be a sufficient alternative to an ivory import ban; legislative controls and quantitative restrictions would also be necessary. In particular, any possible loopholes that *may* allow illegal ivory or ivory imports without proper documentation to enter Japan if it is only slightly worked or cut would need to be closed. In turn it may also be necessary to impose taxes on all worked ivory permanently entering Japan. Moreover, a major advantage of combining taxes with existing and improved controls is to counteract any smuggling of raw ivory in Japan if this exists (Caldwell, 1987). A sales or value-added tax on ivory consumed within Japan would further discourage demand whether the raw ivory was legally imported or smuggled.

---

[10] The above results would suggest that subsequent models should be developed in order to specify a more elaborate price expectations formulation process, drawing on the literature on the economics of demand for storable raw materials as outlined in Williams and Wright (1991).

The ultimate aim of taxes and regulations on net imports of raw ivory into countries that are large end-use consumers such as Japan would be to offer a viable alternative to the current trade ban as a means to controlling the global trade and thus hopefully help end the over-exploitation of elephant populations in Africa. Unfortunately, stricter controls and the use of taxes by one or even a few consuming countries may reduce their domestic consumption of raw ivory significantly, but may not necessarily affect the global legal and illegal trade if this is diverted to new entrepots and end-use consumers who have less strict policies and controls. Some evidence exists of such diversions taking place in the late 1980s in response to the stronger controls imposed by Japan and the strengthening of regulations in Hong Kong. Moreover, there is the unknown factor of how a fall in demand of a large end-use consumer such as Japan will affect the trading price of raw ivory. On the other hand, if Japan were to be successful in reducing its domestic consumption of raw ivory substantially through the combination of taxes and regulations, increased pressure through CITES and other organizations could be put on other consuming countries to follow suit. The demonstration of effective controls is preferable to *ad hoc* implementation of untried and possibly dubious controls.

Finally, a price-inelastic demand for ivory is one of the conditions favourable to the formation of an ivory producers' exporting cartel as part of an international 'Ivory Exchange' (Barbier et al., 1990; and Swanson, 1989). If the African exporters can collude on price fixing and quota arrangements, the low price elasticity of demand for Japan – a major consuming country – suggests that revenues for the cartel can be raised by restricting exports of raw ivory. Alternatively, a small import charge of perhaps around 1-3 per cent, could be levied on Japanese imports strictly to raise revenue.[11] The amount raised could be transferred to African countries to finance elephant management and habitat conservation programmes. Such schemes have the advantage that both producers and major end-use consumers such as Japan would have a vested interest in cooperating to establish a controlled trade based on a sustainable yield of ivory and to minimise illegal harvesting and trade.

The potential policy measures for controlling entrepot trade in ivory are less certain. For example, the complex motivation for Hong Kong's demand for raw ivory imports obscures the responsiveness of imports to changes in real prices. For total imports, higher real raw ivory prices may actually spur a booming re-export trade. It is therefore unclear as to whether a tax on gross ivory imports would reduce demand. In any case, the elasticity estimates indicate that total imports are much more responsive to changes in interest rates and exchange rates (Table 5). Even if we were confident that an increase in real import prices would reduce raw ivory imports, a rising exchange rate and falling real interest rates might offset this effect and cause imports to expand. For net imports the price response is more complicated still. Once again this response appears positive in the current period. This may be due to expanding supply in the carving industry pulling in more net imports, or stockpiling of higher valued raw ivory for future re-export. As in the case of gross imports, however, the price responsiveness of net imports is overshadowed by the impacts of interest rates and the exchange rate.

---

[11] Based on the value of 1987 net imports to Japan, such a tax would raise approximately $190,000 to $560,000 per annum, with only negligible impacts on the total amount imported. In most commodity markets, a revenue-raising import levy of between 1-3 per cent would expect to have virtually no volume consequences, as the effects of such a modest surcharge would be absorbed in the 'normal' price fluctuations due to exchange rate changes, changes in government regulations, supply and demand conditions, etc. (NEI, 1989).

The complicated response of raw ivory imports into Hong Kong to changes in real (net or gross) import prices suggests caution in the use of taxation to control imports. Although imports are more responsive to interest rates and the exchange rate, these variables cannot be used to control Hong Kong's raw ivory trade because they are essentially determined by world economic factors and overall macroeconomic policy. This leaves three options in the control of Hong Kong's entrepot ivory trade: 1) further tightening of Hong Kong's CITES-agreed legislation on ivory imports and exports; 2) getting Hong Kong to accept a ban on its ivory trade; and, 3) controlling trade through Hong Kong by reducing demand by the end-use consuming countries it supplies.

As discussed, the first option may divert Hong Kong's entrepot trade to other entrepot countries who have not adopted or have ignored CITES regulations. The estimated income elasticity for Hong Kong's net imports of raw ivory suggests that in the long run its carving industry might shift to a lesser developed country anyway. Certainly if Hong Kong were to join the international ban on ivory trading and prohibit its trade in raw and newly carved ivory, the shift of its trade to other locations not participating in the ban would be expedited. Thus, unless all existing and potential entrepot countries accept the recent CITES ban on ivory trade, unilateral acceptance by Hong Kong may not have that much of a long-term impact on reducing the world ivory trade.

The demand analysis for Hong Kong yielded an income elasticity of around 0.4 for net imports and over 1.5 for total imports (see Table 5). The implication is that, as the Hong Kong economy continues to develop and grow, it will tend to expand its luxury good re-export market in raw ivory relative to its carving industry. As shown in Figure 4, this appears to have happened up to the mid-1980s. In the late 1980s, the effect of the CITES-inspired Hong Kong legislation, but perhaps more importantly the 1985 legislation in its main export market of Japan, appears to have dampened the Hong Kong re-export trade in raw ivory. This would suggest that controlling and reducing the demand in final end-use countries might be an effective way of limited the re-export markets of entrepot countries.

Reducing the end-use demand for raw and newly carved ivory, as well as ivory-based products, in final consuming countries could have a dramatic impact on controlling the trade not just through Hong Kong but through other existing and potential entrepots. As demonstrated in the demand analysis of Japan, effective policies for controlling final consumption demand can be devised. In the long run, this could be the most effective way of controlling world trade, including reducing the incentive for other entrepots to take up trade displaced from Hong Kong. Further analysis is required, however, of the nature of the entrepot trade through Hong Kong and similar countries to understand the implications of this and even the other two options on world trade.

More detailed disaggregate study and analysis is required of the net import demand of raw ivory into Hong Kong and its re-export trade. It is important to understand more clearly the trends over time determining the supply of raw ivory to the worked ivory industry and the stockpiling of ivory for future re-export, or for future use by the carving industry, or simply as a store of wealth. What proportion of the remaining raw ivory meets final consumer demand in Hong Kong or is smuggled abroad also needs to be determined. In the case of re-exports, more detailed analysis needs to be conducted of the demand for exports from Hong Kong by its main foreign markets. In particular, the Hong Kong-Japanese trade in ivory needs further and continual analysis.

6. CONCLUSION

Economic analysis of the demand for raw ivory by Japan and Hong Kong demonstrates that policy interventions to reduce the demand for raw ivory are possible at the final consuming country stage of trade, and perhaps to control entrepot trade in raw ivory. However, reduction of demand is likely to be *easiest* at the end-use rather than at the entrepot level of the trade. The analysis suggests that a combination of import taxes and regulation is one possible policy option worth investigating, particularly if combined with the participation of ivory producer and consumer countries in an internationally regulated 'Ivory Exchange' (Barbier *et al*, 1990; and Swanson, 1989).

From the point of view of an end-use consumer country such as Japan, although a substantial tax that would double the real price of imported raw ivory might appear to be politically onerous, it is certainly less stringent than Japan's recent decision to ban raw ivory imports, which is essentially an infinite tax. Unfortunately, the recent public and media attention paid to the African elephant's decline – and the subsequent decision by CITES to impose an indefinite ban on the ivory trade – may have already eliminated any serious discussion of a workable system of controlled trade involving taxation as a viable policy option. Such a system, whether involving import taxes in consuming countries such as Japan and/or export taxes imposed by African nations, does take time to implement and operate effectively. We may in any case be ten years too late to adopt such a system to prevent the elephant's decline.

To obtain a more complete understanding of the responsiveness of ivory trade to economic incentives, there is a need to supplement the demand analysis of Japan and Hong Kong by extending the work to cover other major consumers and traders of raw ivory. This involves analysing the trends over time in demand for unworked ivory of such countries as the USA, the EEC, Taiwan, and China.

Finally, there is an urgent need to examine the inter-relationship between the raw and worked ivory trade. The trends in the worked ivory industry and their implications for the demand for ivory as a raw material need to be better determined. As indicated in the demand analysis of Hong Kong, this inter-relationship is extremely significant. Further analysis of the demand for worked ivory exports from Hong Kong in its main foreign markets, such as the United States, is also required. Without such an analysis, only a part of the story of the entrepot trade in ivory can be learned, thus inhibiting our ability to develop policy options for controlling the global trade in ivory and ivory products.

APPENDIX

The elasticity estimates for raw ivory demand reported in Tables 4 and 5 are based on log-linear time series regressions. The 'best estimates' corresponding to these models (t-statistics in parentheses) are:

Japan, Model 1

Corrected for first-order serial correlation:

$$\begin{array}{lccc} C = 4.826 & -\ 0.516P & +\ 0.999Y & -\ 1.419D \\ \phantom{C = {}}(1.23) & (-1.73) & (2.68) & (-4.95) \end{array}$$

period: 1952-87 $R^2 = 0.807$ adj $R^2 = 0.782$

$S.E. = 0.324$ $D\text{-}W = 1.97$ $F_{3,32} = 32.5$

Japan, Model 2, with constant

Corrected for second-order serial correlation:

$$\begin{array}{lccccc} C = 4.901 & -\ 0.702P & +\ 1.147Y & -\ 0.342R & -\ 0.397\mathrm{R}(-1) & -\ 1.419D \\ \phantom{C = {}}(0.61) & (-2.23) & (1.90) & (1.36) & (-1.40) & (-3.83) \end{array}$$

period: 1954-87 $R^2 = 0.844$ adj $R^2 = 0.802$

$S.E. = 0.315$ $D\text{-}W = 1.90$ $F_{5,29} = 20.1$

Japan, Model 2, without constant

Corrected for first-order serial correlation:

$$\begin{aligned} C = {} & 0.718P + 1.530Y + 0.372R - 0.257R(-1) - 1.20D \\ & (-2.39) \quad (6.43) \quad (1.55) \quad (-1.10) \quad (-4.03) \end{aligned}$$

period: 1953-87 $R^2 = 0.816$ adj $R^2 = 0.784$
$S.E. = 0.324$ $D\text{-}W = 1.95$ $F_{5,30} = 25.7$

where

C = Japanese consumption (net imports) of raw ivory
P = Real Japanese (net) import price of raw ivory (1980 prices)
Y = Real Japanese gross national product (1980 prices)
D = Dummy variable for 1986
R = Real Japanese discount rates (1980 prices)
$R(-1)$ = R lagged one year.

Hong Kong, Model 1, net imports

Corrected for second-order serial correlation:

$$\begin{aligned} C = {} & 13.982 - 0.514P(-1) + 0.352P(-2) + 4.436Y - 3.519Y(-1) \\ & (10.2) \quad (-5.15) \quad (2.87) \quad (7.39) \quad (-5.72) \\ & - 0.857Y(-2) + 4.653E - 5.711E(-1) - 0.395D \\ & (-1.61) \quad (4.99) \quad (-7.20) \quad (-3.0) \end{aligned}$$

period: 1966-87 $R^2 = 0.934$ adj $R^2 = 0.874$
$S.E. = 0.149$ $D\text{-}W = 2.22$ $F_{8,22} = 15.6$

Hong Kong, Model 1, total imports

Corrected for first-order serial correlation:

$$\begin{aligned} C = {} & 14.283 + 0.421P - 0.549P(-1) + 1.088Y - 0.806Y(-1) \\ & (12.50) \quad (1.67) \quad (-2.38) \quad (1.94) \quad (-1.46) \\ & + 1.123E - 1.409E(-1) - 2.082E(-2) \\ & (-1.39) \quad (-1.16) \quad (-2.22) \end{aligned}$$

period: 1964-87 $R^2 = 0.861$ adj $R^2 = 0.787$
$S.E. = 0.192$ $D\text{-}W = 2.04$ $F_{7,24} = 11.6$

Hong Kong, Model 2, net imports

Corrected for second-order serial correlation:

$$\begin{aligned} C = {} & 7.723 + 0.479P - 0.199P(-2) + 0.398Y + 0.490R \\ & (3.20) \quad (10.44) \quad (-4.13) \quad (2.82) \quad (2.43) \end{aligned}$$

38

UMENT DELIVERY

WHARTON REPROGRAPHICS

ELSEVIER

Ecological Economics 19 (1996) 147–155

ECOLOGICAL ECONOMICS

ANALYSIS

The ivory trade ban: Is it effective?

Jyoti Khanna [a,*], Jon Harford [b]

[a] Department of Economics, Colgate University, Hamilton, NY 13346, USA
[b] Department of Economics, Cleveland State University, Cleveland, OH 44115, USA

Received 25 July 1995; accepted 27 March 1996

Abstract

This paper examines the effectiveness of the ivory trade ban when producer and consumer states have different incentives to comply with such a ban. The paper distinguishes between two types of producer states: those with stable elephant population and those with declining elephant population. We compare the independent enforcement level of each of these producer states and a consumer state and show the suboptimality of these levels compared to optimal enforcement level that would result from a cooperative effort. The costliness and imperfectness of enforcing constraints on the private taking of elephants creates intergovernmental externalities. Thus, in the absence of binding international law, regulation would have to be supplemented with incentives.

Keywords: Conservation; Ivory trade; Suboptimality; Enforcement; Incentives

1. Introduction

Over the past few years, conservation efforts to save the African elephant have moved beyond the boundaries of their natural habitat to encompass the entire world. In 1989, the African elephant was placed on Appendix I of the convention for international trade in endangered species (CITES), thereby banning trade in ivory and other elephant products among the signatory nations. The ivory trade ban, in its first few years, was hailed as a success by western nations and conservation groups. Now, five years later, feelings about the ban are mixed; poaching, that had initially declined, has started its climb back to the pre-ban levels, and African nations, burdened with the responsibility to enforce the ban without much help from the most ardent proponents of the ban in the west, are increasingly feeling the financial strain of such a conservation strategy. This paper looks at the effectiveness of a trade ban as a conservation tool, given the costliness and the imperfectness of enforcing such constraints on the world community.

Elephants, like other wild species, are harvested legally and illegally for their high commercial value. In the absence of any well-defined or enforceable property rights, the exploitation of wild species is considered to be far greater than that of other natural resources of a country (Johnson, 1990). Although state ownership exists, there are few, if any, constraints on harvesting. This, combined with the large

---

[*] Corresponding author. Tel.: (1-315) 824-7527; Fax: (1-315) 824-7726; e-mail: jkhanna@center.colgate.edu

0921-8009/96/$15.00 Copyright © 1996 Elsevier Science B.V. All rights reserved.
PII S0921-8009(96)00043-2

home range (20 km^2 to 3000 km^2) that elephants inhabit, implies an open-access resource over which no one has any effective control.

Countries, which are theoretically vested with the ownership of this resource are faced with two conflicting objectives: Maximize revenues from the consumptive use of this resource while at the same time maintain a stable population to maximize the non-consumptive use and to protect the ecosystem. The removal of keystone species, like the elephant, can trigger an avalanche of extinctions and lead to a collapse of the eco-system; at the same time, however, products from these species command a high price in the world markets. In 1989, for instance, elephant ivory attracted a price of $140 per pound. Given the open-access nature of this resource, however, taking of elephants was not confined to government-sponsored harvesting but included illegal taking as well. Elephants were illegally killed at the rate of three a day. Each dead elephant could yield about $3600 for the middleman at time when the average wage of a worker was no more than $1000 a year (Western, 1993). Ivory poaching is cited as the main reason why the Africa-wide elephant population was cut in half between 1979 and 1989. In Kenya alone the elephant population plunged from 167 000 in 1973 to a mere 16 000 by 1989. With this rate of harvesting, it was estimated that elephants could be extinct by the year 2010. A similar trend holds for other wild species as well. The population of the black rhinoceros, for instance, has declined by 95% since 1970 (New York Times, 1995).

Such alarming trends created a public outcry in the 1980s when it was increasingly felt that no state had the right to allow destruction of a species even if its habitat was restricted to that country's own territory. The result was the emergence of demand-oriented conservation policies developed by the world community. The convention on international trade in endangered species (CITES) came into effect, a treaty regulating trade in imperiled plants and animals. This treaty places endangered species on Appendix I which means member nations cannot trade with one another in the products of these species. Thus, over the last decade, all commercial trade in leopard skin, elephant ivory and the rhinoceros horn has been regulated or banned by an international agreement under CITES.

The 'embargo' approach to conservation implied by a ban on all legal trade in the products of these species depends for its success on complete compliance by all producer and consumer states. International agreements, however, are not mandatory and even if accepted, can hardly be enforced. States can take a reservation from the agreement thereby reducing the effectiveness of the ban. Partial enforcement allows some states to carry on trade in the products of species listed on the CITES charter. The possibility of trade creates externalities for states complying with the ban. A potential market for the products from these species increases the incidence of poaching in compliant states thereby imposing real costs on the government. This is particularly true for species with grazing area spread over several nation-states, making policing even more difficult. The effect of a partial international ban, therefore, is for legal trade to provide a camouflage for illegal trade. With the trade paths for legal trade well in place, an increased volume on the illegal circuit would not entail any additional costs or time. The overall effect, therefore, would be an increase in the (illegal) killing of these animals rendering the conservation efforts useless.

This paper analyzes the incentives of producer and consumer states to comply with the international ban on trade in wild species with a commercial value. Taking the example of the African elephant, we determine the enforcement levels of two producer states, with different objective functions, to stop the illegal flow of ivory. We compare the Nash–Cournot adjustment levels to the Pareto-optimal levels to identify the direction of suboptimality associated with the independent adjustment levels. The plan of this paper is as follows: Section 2 looks at ivory trade to understand the appropriateness of a trade ban, Section 3 gives the independent adjustment levels of enforcement against illegal ivory, Section 4 gives the Pareto-optimal levels of enforcement and concluding remarks are contained in Section 5.

2. The trade ban

The absence of definable property rights for wild populations has led the world community to adopt trade restrictions to protect them from extinction.

The convention on international trade in endangered species (CITES) is designed to serve this purpose. The treaty was signed in 1973 and by 1989 there were 103 countries party to this agreement. CITES establishes a three-tiered structure in which levels of protection are a function of the degree of threat to the survival of the species. Appendix I provides the highest level of protection by banning all trade in the products of the species. Appendix II provides an intermediate level of protection by regulating rather than banning trade. Thus, a limited amount of trade, approved either by a central authority or by the individual governments, is allowed in the species or products thereof. Appendix III offers the least protection since it involves a mere identification of potentially threatened species. States may or may not impose any trade restrictions on species listed on Appendix III (Barbier et al., 1989).

Beginning in 1978, the African elephant was identified as a threatened species and listed on Appendix II of the CITES agreement. Due to rising demand for ivory during the 1980s, however, the price of ivory soared and poaching accelerated. There was growing demand for updating the elephant to Appendix I which would ban all trade in ivory. Kenya, under the leadership of President Daniel Arap Moi, declared an all out war to save the elephant. Poachers were to be shot on sight and, in the same spirit, President Moi set fire to 2500 ivory tusks, worth $3 million. It was felt that the world was losing the elephant and an increasing pressure was mounted by the western nations and conservation groups to ban all trade in ivory. In June 1989, France, USA, West Germany and the European Community imposed a moratorium on all ivory imports. In October 1989, the African elephant was officially moved to Appendix I. Five southern African states of Zimbabwe, Botswana, Namibia, Malawi and South Africa, however, did not sign the treaty. The first four of these southern states established the South African Center for Ivory Marketing (SACIM) as the exporting agency of ivory from its members. South Africa has its own control and marketing system similar to SACIM.

The embargo model as envisaged by the Appendix I listing is derived from the practical impossibilities inherent in distinguishing illegal ivory from legal ivory. Its proponents reasoned that the more legal ivory there is, the easier it would be to place the illegal ivory into the trade stream (80% of ivory traded during the 1980s was illegal, according to a report published by the African Elephant Conservation Coordinating Group (1989)). A total ban, therefore, was seen as the only way to save the elephant. The supporters argued that any resumption of trade would revitalize the market and put the poachers back in business. The management model, as conceived by the five states who took reservation from the treaty, attempted to limit and control ivory. Sale of ivory was to be permitted from confiscated poached ivory, culling operations and from elephants that die a natural death. The arguments were that revenues from these sales would provide for stricter enforcement. A ban, it was argued, would simply drive the price of ivory upwards and create more incentives for the poachers (Harlan, 1990). In Zimbabwe, for example, loss of ivory revenues led to a scaling back of anti-poaching operations which resulted in 100 illegally killed elephants in 1990 as opposed to only 10 illegal deaths a year earlier. It can cost a country millions of dollars just to police elephants, and with no revenues coming from ivory and sport fees, it was felt that conservation programs would greatly suffer (Glennon, 1990).

Each of the two enforcement models is defined in terms of the shortcomings of the other. The validity of the two models then, it would seem, is an empirical question. An analysis of the incentives of the different players, however, can provide an answer to how a ban on trade in products of these species will affect the enforcement of the ban in different countries. To determine the appropriateness of each model to preserve the population of elephants, however, it is important to understand ivory trade. Until recently, little was known of ivory trade. In 1988, the World Wildlife Fund sponsored a meeting in Lusaka, Zambia, to explore how the elephant could be best protected. The consensus that emerged was that not enough was known about ivory trade to make any recommendation (Glennon, 1990). Since then, however, broad contours of trade have emerged. The principal countries of origin are Kenya, Tanzania, Botswana, Zambia, Zimbabwe and South Africa. There are two types of consumption centers: the carving centers which consume raw ivory and the final consumers of worked ivory. For the most part, raw ivory from entrepots like Singapore, South

Africa, Burundi and United Arab Emirates, moves to the carving centers of the world like Hong Kong, Japan and Dubai. The worked ivory is then either sold domestically, as in Japan, or shipped to final consumers like the USA and Europe. The USA consumed 20% of all worked ivory that entered the international trading system in any given year; Japan consumed about 40%; about 25% of total volume is consumed in Asia where there is a strong latent demand. Since the African elephant was added to Appendix I in 1989, however, the USA, Japan and Europe have banned all imports of ivory.

To protect the elephant stock, an ivory trade ban must create incentives for agents to comply with the ban. These incentives should be such that they generate acceptance and implementation of the scheme by all the agents involved in ivory trade: the producing, manufacturing (carving centers) and the consuming countries of the world. The interests of these agents, however, differ within each group and across groups. Within the producing group, for instance, Kenya, whose elephant stock is valued much more highly for its tourist-generating capacity than for its ivory revenues, has been at the forefront of banning all trade in ivory. In 1988 tourism was Kenya's top foreign exchange earner, surpassing coffee and tea. Much of this tourism was related to wildlife. Kenya's wildlife-related tourist revenues are estimated at $350 million per year, with an elephant's tourist-related worth estimated at $14 375 per year (Olindo, 1991). Other producers like Zimbabwe, Malawi, Botswana and Mozambique have relatively small wildlife-related tourist revenues. In Malawi these revenues are estimated at $1 million per year and in Zimbabwe at $60 million per year.

Aside from the importance of tourist revenue generating capacity of elephant populations, producers differ in the size and management of their elephant populations. Kenya, for example, accounts for 18% of East African elephant population and has designated 12% of its land mass as protected area. About 64% of Kenya's elephant population resides in these protected areas with only 7.5% getting a 'secure' level of protection. The trend of populations in these protected areas, however, is dismaying. For 62% of the protected areas the trend was downward; in only 2% was the trend indicated as being upward. As a result, Kenya lost 46% of its elephant population over a period of 6 years between 1981 and 1987.

Some southern African countries like Zimbabwe, Namibia, Botswana, Malawi and South Africa, on the other hand, show stable or upward trends in their elephant populations. Zimbabwe accounts for 24% of its regions elephant population and has designated 85% of its land mass as being protected. About 95% of Zimbabwe's elephant population resides in these protected areas. The trend in 89% of the protected areas was characterized as being upward. Thus, Zimbabwe's elephant population declined by only 12% in the 6 year period between 1981 and 1987. These trends are even more significant for Botswana and South Africa.

The producer countries, therefore, based on their consumptive or non-consumptive use of the resource, have taken an opposing stand on the trade ban on ivory. Countries with rapidly declining populations and with considerable tourist revenue losses, supported the ban on ivory trade hoping this would provide a respite for their elephant populations. Countries with stable populations, on the other hand, opposed the trade ban on the grounds that it would reduce their revenues from trade in a resource that is well managed. The voting record of nations on the ban highlights this fact. Elephant population had decreased by 74% between 1979 and 1989 in nations voting for the ban, while the population had increased by 9% during the same period in nations voting against the ban (Kreuter and Simmons, 1995). The manufacturing nations also have a large vested interest in the continuation of ivory trade. Predictably, all manufacturing countries except Japan took a reservation from the trade ban.

The final consumers of ivory can be divided into two groups: first, the traditionally large importers like the US, the EEC and Japan who have decided to comply with the ban and, in some cases, the ban has been initiated in these countries; second, the emerging consumers in the Middle East and Asia who are non-CITES countries and are not under any obligation to limit their imports of ivory. Thus, there exist incentives and pressure on agents in the producing countries to satisfy this demand for ivory. Moreover, even in countries complying with the ban, there will always exist private incentives for smuggling the product into the domestic market. A ban imposed by a consuming nation does not imply that all consumers have adopted the ban personally. In Japan, for example, the price of (illegal) ivory has doubled since the ivory ban went into effect suggesting there is an underground market for ivory. In 1993, the price for unworked ivory in Hong Kong was quoted at \$2000 per kg and for worked ivory at \$20 000 (Kreuter and Simmons, 1995). Prices in the USA, on the other hand, have fallen precipitously indicating a genuine decrease in demand. Effectiveness of a ban on imports of certain goods, therefore, requires acceptance on part of the consumers, failing which the government has to commit resources to monitoring and enforcement of the ban. There exists, therefore, little incentive on part of the consumer states to commit resources for the purpose of restricting trade, even if these countries attach an existence value to elephants. Japan has been a reluctant member of the agreement to ban trade in species and even now ranks the lowest in protection of wild land species. Thus it appears that various agents engaged in ivory trade have made decisions consistent with their individual economic objectives. These objectives are not complementary, but competing across agents and a mere statement of restriction of trade will not alter or stop the trade flows.

3. Nash–Cournot enforcement levels

We assume there are two producer countries denoted by Z and K. Country Z attaches a relatively low value to the non-consumptive use of its resource (perhaps because of a stable elephant population). Country K, in contrast, places a high value on the non-consumptive use of its elephant stock. Predictably, country K supports the ban on ivory trade whereas country Z takes a reservation from the embargo. The consuming nations are represented by a single country J which imports ivory, but also places an existence value on the elephant stock, where existence value is broadly interpreted to mean any value associated with the enjoyment of living elephants. We can define the net benefits for each country in terms of revenues from consumptive and non-consumptive use of the elephant stock less the enforcement costs incurred to comply with the ban or to regulate the illegal flow of ivory.

Let country K's benefit function be defined as follows:

$$B_K = B_K\{T_K[E_K\{P_K(N_K, N_J, I_Z)\}]\},$$

where B_K = benefits of country K, T_K = amount of tourism in country K, E_K = elephant stock in country K, P_K = amount of poaching in country K, N_i = level of enforcement by country i, $i = K, J$, I_Z = amount of ivory traded by country Z.

According to the above formulation, country K's benefits from the elephant stock obtain solely from tourist revenues. The stock of elephants, in turn, depends on the level of poaching in that country. Further, poaching in country K depends on the enforcement effort in country K as well as enforcement effort in the consuming country J. Since country Z does not comply with the ban, its enforcement effort is not directed at reducing the inflow of poached ivory from country K. However, legal ivory trade in country Z increases the amount of poaching in country K by creating opportunities for agents to sell illegal ivory through the legal conduits. Banning trade without adequate enforcement structure provides incentives for expanding the illegal component of trade.

Independent adjustment requires that each country choose its own enforcement level, given the equilibrium adjustment levels of the other countries. The maximization problem of country K, therefore, can be represented as

$$\Psi = B_K\{T_K[E_K\{P_K(N_K, N_J, I_Z)\}]\} - \vartheta_K N_K, \quad (1)$$

where ϑ_K is the per unit enforcement cost in country K. The first-order condition (FOC) associated with (1) is

$$\frac{\partial B_K}{\partial T} \cdot \frac{\partial T}{\partial E} \cdot \frac{\partial E}{\partial P_K} \cdot \frac{\partial P_K}{\partial N_K} - \vartheta_K = 0. \quad (2)$$

According to Eq. (2), country K equates the marginal benefit from higher tourist revenues to the marginal cost of enforcement, given the equilibrium levels of N_J and I_Z. Let the resultant Nash equilibrium level be denoted by $\hat{N}_K$.

The benefit function for the second producer Z can be defined in the same way as country K. However, benefits for country Z derive only from the sale of ivory. There is no non-consumptive value placed on the elephant stock. The maximization problem of country Z may be written as

$$\Psi_Z = B_Z\left[I_Z, P_K(I_Z, N_J), P_Z(N_Z, N_J)\right] - \vartheta_Z N_Z, \tag{3}$$

where B, I, P and N, as defined above, now pertain to country Z. The benefit function of country Z depends on ivory sales and the amount of poaching in its own country and in country K. Poaching in country K increases the benefits for country Z since poached ivory can reach the world market only through country Z, thereby bringing benefits associated with a larger volume of traded ivory. Higher poaching in country Z, however, reduces benefits for Z since local poaching may be done by poachers from other countries. For example, the former President of Somalia, Mohamed Siyaad Barre, explicitly authorized the import into Somalia of raw ivory from neighboring countries (UNEP Report, 1989) In any case, poaching reduces the legal harvest of the government. Hence, enforcement in country Z is directed at controlling the illegal harvest of ivory in country Z. Country Z, therefore, faces two decision variables; I_Z, the level of ivory to be traded and N_Z, the level of enforcement. The FOCs associated with Eq. (3) are

$$\frac{\partial B_Z}{\partial I_Z} + \frac{\partial B_Z}{\partial P_K} \cdot \frac{\partial P_K}{\partial I_Z} = 0 \tag{4}$$

$$\frac{\partial B_Z}{\partial P_Z} \cdot \frac{\partial P_Z}{\partial N_Z} - \upsilon_Z = 0. \tag{5}$$

Eq. (4) determines the equilibrium amount of ivory traded, $\hat{I}_Z$; with no harvesting costs, country Z will export ivory until the marginal benefit from legal and illegal harvest goes to zero. Eq. (5) gives the independent enforcement level of country Z. Equilibrium enforcement level, $\hat{N}_Z$, is such that the marginal benefit from reduced poaching is equal to the cost of enforcement.

Consuming country J's benefits are derived from the consumption and non-consumption uses of the elephant stock. That is

$$B_J = B_J(I_J, E_J),$$

where $I_J = I_Z + P_K + P_Z$ and $E_J = E_K + E_Z$. To preserve the elephant stock country J incurs enforcement costs to stop the flow of poached ivory from countries K and Z. The maximization problem is:

$$\Psi_J = B_J\left[(I_Z + P_K + P_Z), (E_K + E_Z)\right] - \vartheta_J N_J \tag{6}$$

The FOC associated with (6) is

$$\frac{\partial B_J}{\partial I_J} \cdot \frac{\partial P_K}{\partial N_J} + \frac{\partial B_J}{\partial I_J} \cdot \frac{\partial P_Z}{\partial N_J} + \frac{\partial B_J}{\partial E_J} \cdot \frac{\partial E_K}{\partial P_K} \cdot \frac{\partial P_K}{\partial N_J} + \frac{\partial B_J}{\partial E_J} \cdot \frac{\partial E_Z}{\partial P_Z} \cdot \frac{\partial P_Z}{\partial N_J} - \vartheta_J = 0. \tag{7}$$

Eq. (7) shows the equilibrium the enforcement level such that the net marginal benefits from a larger elephant stock are equated to the marginal cost of enforcement in country J.

Eqs. (2), (5) and (7) above determine the Nash–Cournot enforcement levels in each country. Given that country Z through ivory trade imposes an externality on country K and country J through enforcement imposes an externality on country K, the independent adjustment levels are likely to be suboptimal. In the next section we derive the socially optimum enforcement levels for each country and compare them to the independent adjustment levels.

4. Optimum enforcement levels

Socially optimum enforcement levels require agents to internalize the externalities. This could result from a cooperative interaction between agents where the optimum level of aggregate enforcement is determined jointly by the agents, and given that, each agent determines the individual level of enforcement. In the absence of any cooperative solution the optimum levels could be enforced by a central authority, in this case the CITES Secretariat, on the individual states. A social optimum requires that total benefits from the elephant stock and ivory revenues be maximized, given the costs of enforcement. Thus, each country's enforcement level would be such that total benefits are maximized. The Lagrangian associated with the social optimum is the sum of the benefits of each agent, given the total costs. That is,

$$\Psi_T = \Psi_K + \Psi_Z + \Psi_J,$$

where Ψ_T is the Lagrangian associated with the world optimum level. The individual socially optimum enforcement level is given by the equilibrium associated with $\partial\Psi_T/\partial N_i$, $i = K, Z, J$. The necessary condition for a social optimum with respect to country K's enforcement is given by

$$\frac{\partial B_Z}{\partial P_K}\cdot\frac{\partial P_K}{\partial N_K}+\frac{\partial B_J}{\partial I_J}\cdot\frac{\partial P_K}{\partial N_K}+\frac{\partial B_J}{\partial E_J}\cdot\frac{\partial E_K}{\partial P_K}\cdot\frac{\partial P_K}{\partial N_K}+\frac{\partial\Psi_K}{\partial N_K}=0. \tag{8}$$

To attach any meaning to the equilibrium level N_K^* associated with Eq. (8) we need to sign the partials. The first term in Eq. (8) is clearly positive since higher enforcement in country K reduces poaching in country K and thereby reduces revenues from poached ivory for country Z. The second term, $\partial B_J/\partial I_J\cdot\partial P_K/\partial N_K$, captures the effect on country J of lower ivory imports and is, therefore, negative. A larger elephant stock, however, increases the marginal benefits for country J and this is given by the term $\partial B_J/\partial E_J\cdot\partial E_K/\partial P_K\cdot\partial P_K/\partial P_k/\partial N_K$. The last term in Eq. (8) is the independent adjustment term for country K. Evaluating Eq. (8) at the Nash–Cournot equilibrium levels of enforcement in all the three countries ($\hat{N}_K$, $\hat{N}_Z$ and $\hat{N}_J$), will allow us to compare the two enforcement levels and determine the direction of suboptimality associated with independent maximization. Thus,

$$\frac{\partial\Psi_K}{\partial N_K|\hat{N}_K}>0,$$

if

$$\frac{\partial B_J}{\partial E_J}\cdot\frac{\partial E_K}{\partial P_K}\cdot\frac{\partial P_K}{\partial N_K}>\frac{\partial B_Z}{\partial P_K}\cdot\frac{\partial P_K}{\partial N_K}+\frac{\partial B_J}{\partial I_J}\cdot\frac{\partial P_K}{\partial N_K}.$$

That is, if the marginal benefit from tourism and existence value for the consuming country J from a larger elephant stock is greater than the sum of the loss in marginal benefit from lower (poached) ivory flows to country Z and the consumers in country J, then, assuming a strictly concave benefit function, we get,

$$N_K^* > \hat{N}_K$$

That is, the Nash–Cournot enforcement level ($\hat{N}_K$) is less than the Pareto-optimal level N_K^*. Thus, for the embargo model to achieve its objective of conservation of endangered species requires countries that place value on living elephants to compensate the producers that comply with the ban to induce them to increase their enforcement.

The socially optimum enforcement level to the producing country that does not place any existence value on the elephant stock is given by the equilibrium level associated with the following:

$$\frac{\partial B_J}{\partial I_J}\cdot\frac{\partial P_Z}{\partial N_Z}+\frac{\partial B_J}{\partial E_J}\cdot\frac{\partial E_Z}{\partial P_Z}\cdot\frac{\partial P_Z}{\partial N_Z}+\frac{\partial\Psi_Z}{\partial N_Z}=0 \tag{9}$$

Let the equilibrium level associated with (9) be denoted by N_Z^*. The first term in Eq. (9) is the loss in marginal benefit to consumers from lower levels of (poached) ivory flows due to higher enforcement in country Z. The second term captures the increase in marginal benefits to consumers from a larger elephant stock. A comparison of the Pareto-optimal enforcement level with the Nash–Cournot level shows that the independent adjustment level is suboptimal if the existence value is greater than the (marginal) loss in consumer surplus. That is, to achieve higher enforcement, consumers with a high existence value have to compensate the loss in consumer surplus from lower ivory flows. Thus,

$$\frac{\partial\Psi_T}{\partial N_Z|N_Z}>0$$

The independent enforcement level in country Z is suboptimal because it does not take into account the existence value to consumers of its elephant stock. A threat to stop consuming poached ivory or ivory without proper certification will increase the enforcement level in country Z.

The socially optimum level of ivory exports is given by $\partial\Psi_T/\partial I_Z$. That is,

$$\frac{\partial B_K}{\partial T_K}\cdot\frac{\partial T_K}{\partial E_K}\frac{\partial P_K}{\partial I_Z}+\frac{\partial B_J}{\partial I_J}+\frac{\partial B_J}{\partial I_J}\cdot\frac{\partial I_J}{\partial P_K}\frac{\partial P_K}{\partial I_Z}+\frac{\partial B_J}{\partial E_J}\cdot\frac{\partial E_K}{\partial E_K}\frac{\partial P_K}{\partial I_Z}+\frac{\partial \Psi_Z}{\partial I_Z}=0. \tag{10}$$

The first term in Eq. (10) is the loss in the marginal benefits to country K due to poaching in its country brought about by trade opportunities available in country Z. The second and the third terms in the equation are positive and represent the increase in marginal benefits to consumers in country J due to higher ivory flows. The fourth term in the equation represents the loss in marginal benefits to the non-consumption users of this resource. The last term is the individual adjustment problem of country Z, given earlier by Eq. (4). Evaluated at the Nash–Cournot equilibrium level, $\hat{I}_Z$, the last term drops out. The Nash–Cournot equilibrium amount of ivory traded will be suboptimal if,

$$\frac{\partial B_J}{\partial I_J}+\frac{\partial B_J}{\partial I_J}\cdot\frac{\partial I_J}{\partial P_K}\cdot\frac{\partial P_K}{\partial I_Z} > \frac{\partial B_K}{\partial T_K}\cdot\frac{\partial P_K}{\partial I_Z}+\frac{\partial B_J}{\partial E_J}\cdot\frac{\partial P_K}{\partial I_Z}.$$

That is, if the marginal benefit to consumers from ivory sales is greater than the loss in marginal benefit of country K from a lower elephant stock, then

$$I_Z^* > \hat{I}_Z.$$

If $\partial B_J/\partial I_J = 0$, then $\partial\Psi_T/\partial I_Z|\hat{I}_Z < 0$, implying that with a genuine decline in demand for ivory, the amount of ivory traded under independent adjustment is too high compared to the Pareto optimal level. This has been the case with ivory over the past couple of years. With attitudes to ivory-buying having changed dramatically, there seems to be little market for this product. For the rhino case, however, where the horn is bought for medicinal purposes, demand hardly seems to have been affected by the new wave of conservationist attitudes (The Economist, 1993).

Finally, enforcement in consuming country J affects the amount of ivory traded and, hence, the conservation efforts. Higher enforcement in the consuming country has an effect on poaching activity in the producer countries. The Pareto-optimal enforcement in the consuming country is given by the equilibrium level associated with the following expression:

$$\frac{\partial B_K}{\partial T_K}\cdot\frac{\partial E_K}{\partial P_K}\cdot\frac{\partial P_K}{\partial N_J}+\frac{\partial B_Z}{\partial P_K}\cdot\frac{\partial P_K}{\partial N_J}+\frac{\partial B_Z}{\partial P_Z}\cdot\frac{\partial P_Z}{\partial N_J}+\frac{\partial \Psi_J}{\partial N_J}=0, \tag{11}$$

where the terms, partial $B_K/\partial T_K \cdot \partial E_K/\partial P_K \cdot \partial P_K/\partial N_J$ and $\partial B_Z/\partial P_Z \cdot \partial P_Z/\partial N_J$ are positive and denote the increase in marginal benefits for country K and Z from lesser poaching. The term, partial $B_Z/\partial P_K \cdot \partial P_K/\partial N_J$ gives the loss in marginal benefits for country Z from lesser poaching in country K. The reduced level of poaching in both the producer countries is the result of greater enforcement in the consuming country. Evaluating Eq. (11) at the Nash–Cournot levels will give

$$N_J^* > \hat{N}_J,$$

when the increase in marginal benefits for country K and Z from higher enforcement in country J is greater than the loss in marginal benefits for country Z from lower levels of trade. Efforts by a central authority to certify legal imports will produce an automatic movement toward optimal level of enforcement in J since poaching would become unprofitable.

5. Conclusions

This paper has looked at the aspects of the externality problem that permeate the efficient use of elephants. The sustainable utilization of elephants as a source of ivory requires that the open access 'tragedy of the commons' problem be overcome by individual countries. The costliness and the imperfectness of enforcing constraints on the private taking of elephants for ivory has created further intergovernmental externalities. This is because of the potential for legal ivory to provide a cover for illegal ivory. Certainly Kenya has felt that the legal trade by some other countries has hampered their enforcement efforts. Furthermore, countries allowing legal ivory sales have gained from ivory poached in non-trading states. The ideal system of international regulation would be one that maximizes the present value of net benefits from these species-both living and dead—which presumably implies a relatively stable population of elephants.

In the absence of binding international law, regulation has to be supplemented by incentives. Consumers, who place an existence value on these species without travelling to these regions are essentially free riders. Hence, to preserve these species there has to be a mechanism of payments from the consumers to the 'producer' nations to increase enforcement efforts against illegal killings. African nations, burdened with enforcement expenditures and loss of revenues from ivory sales, are ill-equipped to support this conservation strategy for long. Public awareness against the use of the products from these species can kill demand and eventually stop the unnecessary killing of these animals since poaching would become unprofitable. However, without a genuine decline in demand, as seems to be the case, a trade ban has the effect of pushing trade underground and worsening the prospects for a successful conservation program. Even though the last few years have seen a growing consciousness against ivory use, most intermediate consumers are taking a 'wait and see' attitude and are ready to resume trade as soon as there is a slight increase in demand.

Acknowledgements

We would like to thank Isebill Gruhn and Roger Perman for helpful comments on an earlier version of this paper.

References

African Elephant Conservation Coordinating Group, 1989. African Elephant Action Plan. WWF International, Switzerland.

Barbier, E.B., Burgess, J.C., Swanson, T. and Pearce, D.W., 1989. Elephants, Economics and Ivory. Earthscan Publications, London.

The Economist, 1993. Save the Rhino. 9 October 1993.

Glennon, M.J., 1990. Has international law failed the elephant? Am. J. Int. Law, 84: 65–72.

Harlan, D., 1990. Jumping on the 'ban' wagon: Efforts to save the African elephant. Fletcher Forum World Affairs, 14: 134–162.

Johnson, R., 1990. Commercial wild species rearing: Competing groups and regulation. J. Environ. Econ. Manage., 19.

Kreuter, U.P. and Simmons, R., 1995. Who owns the elephants? The political economy of saving the African elephant. In: ed. T. Anderson and P. Hill, Wildlife In The Marketplace. Rowman & Littlefield, MD.

Olindo, P., 1991. The old man of nature tourism: Kenya. In: ed. T. Whelan, Nature Tourism. Island Press, Washington, DC.

New York Times, 1995. Of rhinos, elephants and humans.

United Nations Environment Program, 1989. The African Elephant. UNEP Report, 30.

Western David, 1993. The balance of nature. Wildl. Conserv., 96: 52–55.

39

WHARTON REPROGRAPHICS

Who Voted Against the NAFTA? Trade Unions Versus Free Trade

John A.C. Conybeare and Mark Zinkula

1. INTRODUCTION

AFTER Representative Tom Sawyer (D — Ohio) voted for the North American Free Trade Agreement on 17 November, 1993, the president of the local United Auto Workers union publicly vowed revenge at the next election (Wines, 1993). The hard fought vote on the NAFTA in Congress provides an opportunity to document the influence on Congressional representatives of the increasingly vehement opposition of labour (and especially organised union labour) to free trade. It also provides another observation for the long debate on the extent to which representatives' voting behaviour reflects constituency economic interests or some other influence, such as ideology or party loyalty.

We begin by briefly summarising the history of labour's stance on trade issues and the reasons why it would rationally oppose the lowering of trade barriers. Relevant aspects of the NAFTA debate are explained and the final vote modelled with a two stage logistic regression. The results suggest that representatives were indeed more likely to vote against the NAFTA if their state contained significant labour opposition to free trade. When these economic interests are held constant, Democrats still deserted their president, confirming the continuing shift of the party away from free trade.

2. LABOUR AND FREE TRADE: HISTORY AND THEORY

The conventional wisdom about American political parties and trade is that during the early part of the twentieth century the Democrats came to represent labour and the Republicans capital (e.g., Baldwin, 1985, pp. 33–78; and Magee, 1989, pp. 177–200). Since labour is assumed to be protectionist and capital free trading, the increasingly protectionist preferences of the Democratic party are

---

JOHN A.C. CONYBEARE is from the Department of Political Science, University of Iowa.
MARK ZINKULA is from the Investment Division, Aegon USA, INC.

easily explained.[1] The correlation is not quite that simple, since other factors (e.g., which party controls the executive) sway`party votes, as some observers have noted (e.g., Magee, 1989). Trade liberalisation has been more common under Democratic presidents, even as the Democratic party has become more protectionist than the Republican party (Magee, 1989, p. 182). Nevertheless, the effect of party on trade preferences is strong and predictively powerful.

Why should we expect labour to be against free trade? If factors of production are mobile, the famous Stolper-Samuelson theory provides clear predictions of lobbying over trade policy (Kemp, 1969, pp. 16–19; and Rogowski, 1987). The US can plausibly be characterised as a capital abundant but labour scarce country. This labour scarcity has been increasing (Magee, 1989, p. 185). Assume for simplicity that the US produces two goods, one capital intensive (computers) and one labour intensive (textiles). According to the classical Heckscher-Ohlin theory of trade, the US will have a comparative advantage in (and export) computers because capital is cheap. It will import textiles because labour is relatively scarce and expensive. What the Stolper-Samuelson theory adds to Heckscher-Ohlin is the insight that labour will lobby for protection, even though labour is used in both industries.

This prediction about pressure group politics is easily explicable. A tariff on textiles will raise the price of textiles, causing domestic textile production to expand at the expense of the computer industry. The textile industry will be drawing in both labour and capital as the computer industry contracts and releases both factors. However, since textiles are labour intensive, this industry will be seeking labour at a faster rate than the computer industry is releasing labour; and it will be desiring smaller amounts of capital than the computer industry is shedding. Excess demand for labour and excess supply of capital will cause the rate of return on capital in the US to fall, and wages will rise. Hence the incentive for the scarce factor (labour) to lobby for protection and the abundant factor (capital) to lobby for free trade. Free trade will have the opposite effect, raising the rate of return on capital and lowering wages.

The relative scarcity of labour is greatest for unskilled American labour (Deardorff and Stern, 1979, p. 129). Since the major import competing industries (e.g., textiles, motor vehicles, fruit and vegetables) use much unskilled labour, it would be in the unions representing unskilled workers that we would expect to find the most vehement opposition to free trade.

The critical assumption behind this model of lobbying is that factors are mobile, leading to the prediction that labour and capital will always seek the opposite trade policy, depending on their relative abundance. Factors are mobile in the long term (even if only because capital depreciates and people die!), but if

---

[1] Though as late as 1962, the AFL-CIO supported the Trade Expansion Act, which authorised US participation in GATT talks (see Pastor, 1980, p. 112).

© Blackwell Publishers Ltd. 1996

factors seek their short term interests lobbying might be based on industries rather than factors. Protection will drive up the return to both labour and capital in the textile industry, leading them to lobby together.

Regardless of whether factors are mobile, we expect to find large sections of the US labour force lobbying for protection. Since import competing industries are mostly labour intensive, a disproportionate share of the workforce will be employed in such industries. We would still expect to see most of the labour force preferring protection, even if labour is immobile (and therefore some labour is in export or other industries hurt by protection). Insofar as the factors employed in industries that decline under free trade are immobile, capital in such industries will join labour in campaigning against free trade. The main source of capital immobility is the specialised nature of capital employed in many of the import competing US industries (e.g., steel and motor vehicles). Labour in these industries is also likely to be relatively immobile because it is unskilled — the very category of labour that has the most incentive to seek protection even when labour is mobile.

Thus, as trade has risen as a proportion of US national income, labour in general (and unskilled labour in particular) has become increasingly opposed to free trade, and this imperative has more than offset whatever interest the median Democratic politician has in appealing to the interest of labour as consumers. In certain industries (notably steel, textiles, and until recently, motor vehicles), capital has joined labour.[2] This electoral strategy clearly makes Democrats vulnerable to Republicans who appeal both to broad based consumer interests and to the faster growing industries in which the US does have a comparative advantage. It also narrows the Democratic electoral base geographically to the northeastern parts of the country that have a decreasing share of the population.

3. LABOUR AND THE NAFTA DEBATE

Customs unions and free trade areas rarely result in large short term ('static') effects. Studies of the formation of the European Economic Community concluded that the short term net benefits were in the order of one per cent members' incomes (Heller, 1968, pp. 168–71). The NAFTA will yield static effects on US income and employment that are, in the aggregate, positive but *trivial*. The total increment to US national income will be about three tenths of one per cent (0.003) of GNP, and the total increase in US employment will be

---

[2] Capital in the motor vehicle industry has become much less interested in protection since the early 1980s, partly because of co-production arrangements with Japanese firms, and partly because most of their Japanese competition is now producing within the United States.

© Blackwell Publishers Ltd 1996

about 170,000 by 1995.[3] Such magnitudes are less than the random monthly changes in these indicators — the average monthly change in real GDP during the last quarter of 1993 was 2.5 per cent, and the seasonally adjusted change in employment for February 1994 was +217,000. The vehemence of the protectionist opposition to the NAFTA appears out of proportion to the size of the aggregate effects.

The paradox is easily resolved: the protectionist agenda is not about *net* national benefits but about the redistribution of employment within the United States. We will import more labour intensive goods, such as textiles, fruit and vegetables, but export more land and capital intensive goods, such as grains, machinery, electronic and chemical goods. This redistribution will especially hurt two overlapping categories of labour. One is labour unions that may have been able to hold the wage structure of import competing industries above free market levels, as a result of imperfectly competitive labour markets (e.g., the United Auto Workers in the motor vehicle labour market). The second is *unskilled* labour, the relatively more scarce type of labour in the United States.[4] The NAFTA will exert the most downward pressure on unskilled wages in the United States, as a result of competition from unskilled labour intensive Mexican industries.[5] Not surprisingly, non-union (mostly unskilled) labour in the southern states of the United States has expressed strong, but unorganised, opposition to the NAFTA. Despite their mixed interests in domestic labour markets, union and non-union workers share a common opposition to the NAFTA.

The heaviest lobbying effort against the NAFTA came from the unions affiliated with the AFL-CIO, beginning with its unsuccessful attempt to stop Congressional approval of 'fast track' legislation for the NAFTA in 1991.[6] The United Automobile, Aerospace and Agricultural Implement Workers (UAW) expressed concern over wage differentials between the Mexican and United States auto industries, and the fear that small car production would shift to Mexico. The Amalgamated Clothing and Textile Workers' Union (ACTWU) is concerned also about wage differentials and inadequate adjustment assistance for displaced workers. The Union of Electronic, Electrical, Salaried, Machine and Furniture Workers (UEESMFW) focuses on wage disparities. The United Food

---

[3] Various studies of the economic effects of the NAFTA are summarised in Lustig et al. (1992). Hufbauer and Schott (1993) estimate the employment increase to be 170,000.

[4] It is also possible that some categories of skilled labour may be hurt (if, for example, US factories hire Mexican engineers), but this is much less likely because skilled labour is not as plentiful in Mexico and therefore not as relatively cheap as unskilled labour.

[5] UCLA economist Edward Leamer estimated that unskilled wages could fall by as much as $1000 per year (cited in Lustig et al. 1992, p. 87). We should note that this downward pressure on unskilled wages might be maintained even in the absence of NAFTA, since technical change has had the same effect (see Bhagwati, 1995).

[6] Much of the information cited in the rest of this section has been derived from various newspaper articles and press releases issued by trade unions in the months prior to the NAFTA vote in Congress.

© Blackwell Publishers Ltd. 1996

and Commercial Workers' International Union (UFCWIU) and the International Ladies' Garment Workers Union (ILGWU) believe that the corporations that employ their workers will move to Mexico. The UFCWIU adds that NAFTA will cause 'dislocation' in virtually every industry covered by the union. By June of 1993, of the top 40 Congressional recipients of labour PAC money, 90 per cent had declared their opposition to the NAFTA (Heritage Foundation, 1993). (Support for the NAFTA was much higher amongst governors and senators, since they represent larger constituencies.)

In some cases, union opposition to NAFTA is inconsistent with labour's interests. The construction equipment manufacturer Caterpillar estimated that its export revenues will increase by $45 million as Mexican import tariffs of 10–20 per cent disappear. Yet Caterpillar workers, members of the UAW, opposed the NAFTA (Feder, 1993). It may also not have occurred to the unions that a reduction in Mexican import barriers will bring some jobs back to the United States. In the auto industry, for example, vehicle exports to Mexico have increased along with exports to the United States from US factories in Mexico. In the case of Ford, the net effect has been job creation in the US, mainly because now 80 per cent of the components in Ford's Mexican assembled cars come from the US (*Business Week*, 1994).

Unions have been supported by groups with an indirect interest in their protectionist agenda. Jesse Jackson's 'Rainbow Coalition' opposed the NAFTA, believing that the burden of adjustment will be born largely by unskilled and low wage black and Hispanic workers. Albert Shanker, the leader of the American Federation of Teachers, stated that his organisation opposes the NAFTA because lost jobs may lower local tax revenues and reduce the salaries and bargaining power of his union.

Unions also sought alliances with environmentalists. The potential for linkage was recognised in the main text of the treaty: an admonition not to reduce environmental standards for competitive advantage (Article 1114). Other than one prominent individual, Ross Perot ('United We Stand'), the principal organisation that explicitly linked and espoused both protectionist and environmental concerns was the Citizens Trade Campaign (CTC). This coalition of 47 state organisations drew funds from unions (13 of the 41 supporting organisations, including the UFCWIU) and environmental groups (e.g., Public Citizen, the Sierra Club, and Friends of the Earth). CTC is a loose populist coalition that aggregates grass roots level protests about the NAFTA from unions and environmental groups that are themselves more locally based than national.

Unions raised two environmental issues: the rights and safety of Mexican workers and the threat to United States environmental standards. Since they have not shown interest in the welfare of Mexicans in the past, it might not be cynical to presume that their espousal of these issues partly reflects self interest (forcing costly environmental rules on Mexico), serves a tactical function (linking them to environmental groups) and helps to publicly legitimate their opposition to the NAFTA. Even sympathetic observers (Diamond, 1993) inferred that the objective of unions in raising 'social' concerns was to kill NAFTA rather than to help Mexican workers obtain better conditions. Unions and environmentalists often disagree bitterly on domestic environmental regulations, but imposing such costs on foreign countries is a tactic on which both can agree, even though the long term strategies may be quite different.

Unions affiliated with the AFL-CIO (e.g., Teamsters, UAW, ILGWU) all included in their public statements reference to the failure of the NAFTA to provide for workplace safety of Mexican workers. The labour side agreement (signed in August 1993, along with the environmental side agreement) did little to assuage union opposition, since it addressed only these *stated* (tactical and legitimating) concerns (notably worker safety and child labour in Mexico), rather than their other — perhaps more *real* — concerns (protecting the jobs of their members). The AFL-CIO states that it was not mollified by the labour side agreement because it did not give enough protection to Mexican workers' rights.

Labour was also joined by groups whose pursuit of social reform lead them to cite environmental-protectionist issues. The Congressional Black Caucus stated that it opposed the NAFTA because of the threat to the environment, jobs in general (and low wage, unskilled minority jobs in particular), the lack of assistance for minority businesses to benefit from the NAFTA, and the trade diversion from black populated islands of the Caribbean (*viz*., sugar exports). Jesse Jackson's 'Rainbow Coalition' and Jerry Brown's 'We the People' presented a similar perspective — populist, grass roots level fears that NAFTA means unemployment and pollution. A few religious organisations also joined this agenda, including the United Church of Christ.

Some of the protectionist interests have been bought off, either by provisions in the original agreement, by later side agreements (e.g., tariff 'snapbacks' in the event of import 'surges') or the last minute promises to negotiate understandings after the passage of the bill. The major protective provision in the original NAFTA proposal concerned rules of origin that stipulated how much local content would be required for a good to be subject to the customs benefits of the treaty (e.g., requiring a clothing item to be produced from yarn made in a NAFTA member; or the 50 per cent content rule for automobiles).

The final days before the Congressional vote yielded other concessions to strategically concentrated interests. The Latino Consensus on NAFTA decided to support the treaty in exchange for a North American Development Bank to lend money to communities injured and to finance environmental restoration along the Mexican border. Citrus fruit, vegetable and sugar producers received promises to limit Mexican exports. Wheat producers were tempted with further restrictions on

© Blackwell Publishers Ltd. 1996

Canadian wheat exports, leading the National Association of Wheat Growers to endorse the NAFTA.

The groups that have managed to secure such protection mostly represent interests that are small enough, in terms of the magnitude of the costs of buying them off, to make such a bargain politically possible (e.g., ball bearings, beef, peanut butter, wheat, orange juice, cucumbers). Trade union opposition to the NAFTA could not be bought off because their rejection of free trade is too deep seated in the factor endowment of the US.

These deals were instrumental in producing the margin of victor in the NAFTA vote, by winning over legislators from the key states of Texas, Florida, Michigan and California. Ironically, Texas and California are also the states that have the largest exports to Mexico. The very 'marginality' of support for NAFTA in these large states made the side deals especially critical.

4. TECHNIQUE, DATA, VARIABLES AND RESULTS

The conventional method for determining the influence of particular interests on a Congressional bill is to regress (using logit or probit techniques) the vote record on variables measuring the causal factors of interest. This has been done for a number of Congressional votes on trade bills (e.g., Goughlin, 1985; Tosini and Tower, 1987; McArthur and Marks, 1988; and Conybeare, 1991). We follow this same technique, with a minor correction for a simultaneity problem.

The dependent variable is the roll call vote in Congress (No = 0, Yes = 1) on the NAFTA, 17 November, 1993. In measuring the economic interests that may drive legislators' preferences, we use state rather than district level data for several reasons. District level data is difficult to obtain, and would have to be aggregated by counties, the boundaries of which frequently do not correspond to Congressional district boundaries. Furthermore, many people work in a different district to that in which they live and vote. Finally, there is a considerable amount of bloc voting by states on trade bills. In twenty five states, more than 70 per cent of the representatives from those states voted the same way on NAFTA.[7] Most other studies of roll call votes on trade bills also use state level data.

Since our preliminary interest is in the influence of labour opposition to the NAFTA, the independent variables were selected with this goal in mind. Labour interests in the NAFTA are measured by three variables: UNEMP (per cent of workforce unemployed, by state), UNION (per cent of manufacturing employment unionised, by state), and AFL (the AFL-CIO's rating of the representative). Since an interest group's rating of a politician is itself a result of

---

[7] Tower and Tosini (1987, p. 25) also found bloc voting by state representatives on the textile import quota bill of 1985.

© Blackwell Publishers Ltd. 1996

past voting behaviour, which is in turn a result of characteristics of the constituency, we use a two stage procedure of regressing the representatives' AFL ratings on all other independent variables, and using the residual as the AFL variable, rather than the actual score.[1] Manufacturing employment as a proportion of total state employment (MANUF) is also included, to capture the magnitude of non-union labour interests. The effect of this last variable may be ambiguous, since it would include high skill labour that may benefit from the NAFTA (e.g., chemical and electronic workers).

Several control variables were also introduced. A majority of Democrats voted against the NAFTA (156 to 102 in favour), while Republicans voted in favour (132 to 43 against). PARTY (party of representative, where Democrat = 0 and Republican = 1) is included to measure the influence of party identification, and especially to determine whether the party division is sustained when economic variables are controlled. The percentage of the popular vote received by the representative in the 1992 election (VOTE92) was also included, to test for the influence of electoral marginality. Empirically, electoral safety is associated with faithful representation of constituents' interests (Erikson, 1974; Fiorina, 1974; Thompson and Moncrief, 1988; Cox, 1987, p. 151; and Conybeare, 1991). A plausible explanation is that pleasing voters is what makes a seat safe. We expect representatives in safe seats to be more likely to avoid the risk of offending constituents by voting against the NAFTA. Finally, representatives from states in which exports to Mexico are a high proportion of gross state product (EXMEX) are likely to have voted for the NAFTA.

The results of the logit analysis may be seen in Table 1, and the interpretation is straight forward. The importance of manufacturing industry in a state was, as anticipated, insignificant. Congressmen from states with significant exports to Mexico naturally were more likely to vote for the treaty.[2] Electoral safety had no effect on the vote. Of most interest is the finding that the three measures of labour interests (AFL rating, unemployment and unionisation) produced the expected negative coefficients. Labour interests clearly push representatives away from favouring freer trade.

---

[8] Since the AFL score is a function of how often the representative voted for the stand favoured by the AFL-CIO, it is not independent of the economic indices of constituency interest used here. In order to isolate that part of the AFL score which measures 'ideological' commitment to labour, the residual of regressing the AFL score on the selected economic variables is used in place of the raw AFL score. The implicit causal model being invoked is that past roll call votes (V_{t-1} are a function of economic interests of the electors (E) and some unmeasured ideological preference of the representative (I); the AFL rating (AFL_t) is computed directly from V_{t-1}, so AFL_t is actually a combination of I and E. Hence regressing AFL_t on E yields a residual that is a proxy for I, so we may then estimate V_t as a function of I and E. This form of instrumental variable has been used by many who examine roll call votes (e.g., Kau and Rubin, 1979).

[9] There is no significant correlation between unionisation and exports to Mexico ($r = -0.129$).

© Blackwell Publishers Ltd. 1996

TABLE 1
A Logit Analysis of the Congressional Vote on the NAFTA, 1993

	Coefficient	*Std. Err.*	*t-ratio*	*P>|t|*
AFL	−0.053676	0.007721	−6.952	0.000
UNEMP	−0.148271	0.090853	−1.632	0.103
UNION	−0.015593	0.007409	−2.105	0.036
MANUF	−1.451062	2.547542	−0.570	0.569
PARTY	1.862157	0.254398	7.320	0.000
VOTE92	−0.001440	0.010507	−0.137	0.891
EXMEX	0.622391	0.164126	3.792	0.000
Constant	0.914222	1.021419	0.895	0.371

Number of obs. = 432
$\text{chi}^2(7)$ = 146.25
Prob > chi^2 = 0.000
Log Likelihood = −224.811
Pseudo R^2 = 0.245

Sources: AFL, PARTY (D = 0, R = 1) and the roll call vote (dependent variable, No = 0, Yes = 1) from *Legi-Slate*; UNEMP for 1992 from US Bureau of Labour Statistics (1994); UNION and MANUF for 1992 from US Department of Commerce (1993a); EXMEX (exports to Mexico as a percentage of 1991 gross state product) from US Department of Commerce (1992 and 1993b); VOTE92 from Barone and Ujifusa (1993).

The Republican preference for free trade is striking, even when economic variables are controlled. Taking the mean values of all the other variables as a baseline for predicting the log of the odds ratio, being Republican rather than Democrat increased the probability of a yes vote by 0.446 to 0.842.[10] This supports the findings of Baldwin (1985), McArthur and Marks (1988), Coughlin (1985), Tower and Tosini (1987), and Baldwin (1985) that on trade bills in the 1970s and 1980s Republicans voted for freer trade, controlling for the import sensitivity of their states. The few Republican protectionists, such as Pat Buchanan, have a limited influence on party policy, partly because they are a minority and perhaps also because their protectionism is not a strong core belief, but merely a side effect of their more general commitment to an isolationist foreign policy.

Some observers have noted that trade liberalising Democratic presidents could usually carry their party with them in a roll call vote on a major trade bill. President Kennedy was able to get the Trade Expansion Act of 1962 (authorising the Kennedy Round of GATT negotiations) past the floor of the House with a vote of 256 to 91. President Carter's Trade Act of 1979 (approving the results of the Tokyo Round of GATT talks) passed by 395 to seven. The NAFTA passed in the House by a vote of 234 to 200. It would appear that by the 1990s, a Democratic president can no longer count on party loyalty to pass a trade liberalisation bill. It should also be noted that the preferences of Democratic congressional representatives are independent of the selected economic characteristics of states, and not merely the result of districts with certain economic attributes selecting Democrats to represent them in Congress. The bivariate correlations are very low between PARTY and UNEMP (r = -0.011), UNION (r = -0.036), MANUF (r = -0.051), and EXMEX (r = -0.042). Democrats voted against the NAFTA even when controlling for these proxy measures of the economic interests of their states.

---

[10] If P is the probability of a yes vote and B is the coefficient for PARTY in the logistic equation, then the partial derivative $\partial P/\partial(\text{PARTY}) = P(1-P)B$, where P must be computed for particular values of the all independent variables.

The NAFTA vote is perhaps symptomatic of a narrowing of the Democrat electoral base that has been going on for several decades. Democratic protectionism is driven by two apparently independent factors: adherence to a moribund ideological prescription for economic planning (of which protection is one element) and an electoral dependence on declining interest groups, notably trade unions in industries in which the United States has been losing its comparative advantage.[11] Whether the current Republican control of the House and Senate will encourage a re-examination of this self-destructive strategy remains to be seen. Nevertheless, unions still have considerable power to block trade negotiations, even if only because those politicians most beholden to unions tend to have relatively safe seats.

5. CONCLUSION

Labour interests helped make the NAFTA vote a narrow victory for free trade. A large majority of Democrats could not follow their president's desire to implement the free trade agreement, though they have reasons that go beyond responding to labour's increasing antipathy to free trade. Although AFL-CIO ratings and party identification were important predictors of the roll call, they do not necessarily suggest an important 'ideological' component in the vote, since these variables themselves may be reflecting some aspects of the economic interest of a district other than the ones specified here.

The political implications of this evidence are not encouraging for the prospects of further reducing world trade barriers, particularly non-tariff barriers that are based on environmental, safety or work standard rules. These are now the most favoured tools of protectionists, allowing them to hide behind a purportedly noble objective. Labour unions and other protectionists will surely seek to maintain their alliance with the environmental movement. The next round of world trade talks, under the auspices of the new World Trade Organisation (the successor to GATT), is expected to focus on the environment and labour

---

[11] Irrespective of the industry, trade unions comprise less than 15 per cent of the workforce in any case.

© Blackwell Publishers Ltd. 1996

standards — with strong support from the Clinton Administration (*Economist*, 1994). Since environmentalists believe free trade undermines local environmental rules, and labour unions seek to impose costly labour relations standards on other countries (or simply block free trade by making such demands), the domestic political forces against free trade in the United States maybe exert a critically negative influence on the prospects of successful world trade negotiations in the near future. Even the recently completed Uruguay Round of GATT negotiations is having difficulty in the Congress, as the Commerce Department encourages protectionists to amend the agreement in order to make import restrictions easier to obtain (Passell, 1994).

REFERENCES

Baldwin, R.E. (1985), *The Political Economy of US Import Policy* (Cambridge: MIT Press).

Barone, M. and G. Ujifusa (1993), *Almanac of American Politics, 1994* (Washington, DC: National Journal).

Bhagwati, J. (1995), 'Trade and Wages: Choosing Among Alternative Explanations', *FRBNY Economic Policy Review* (January).

Business Week, (1994) (21 November), 'What Hath NAFTA Wrought?' 48–9.

Conybeare, J. (1991), 'Voting for Protection', *International Organization*, 45, 57–82.

Coughlin, C. (1985), 'Domestic Content Legislation: House Voting and the Economics of Regulation', *Economic Enquiry*, 23, 437–448.

Cox, G. (1987), *The Efficient Secret* (New York: Cambridge University Press).

Deardorff, A.V. and R.M. Stern (1979), 'American Labor's Stake in International Trade', in W. Adams (ed.), *Tariffs, Quotas and Trade: The Politics of Protectionism* (San Francisco: Institute for Contemporary Studies).

Diamond, S.F. (1993), 'US Labor and North American Integration', in R. Grinspun and M. Cameron (eds.), *The Political Economy of North American Free Trade* (New York: St. Martin's Press).

Economist (1994) (16 July), 'A Disquieting New Agenda for Trade', 55–56.

Erikson, R. (1974), 'Is There Such a Thing as a Safe Seat', *Polity*, 8, 623–32.

Feder, B.J. (1993), 'Caterpillar Sees Free Trade Boon', *New York Times* (21 September).

Fiorina, M. (1974), *Representatives, Roll Calls and Constituencies* (Lexington: Lexington Books).

Heller, R. (1968), *International Trade: Theory and Empirical Evidence* (Englewood Cliffs: Prentice Hall).

Heritage Foundation (1993), *Why the Governors Support the NAFTA* (15 June).

Hufbauer, G. and J. Schott (1993), *NAFTA: An Assessment* (Washington DC: Institute for International Economics).

Kau, J. and P. Rubin (1979), 'Self Interest, Ideology and Logrolling in Congressional Voting', *Journal of Law and Economics*, 22, 365–348.

Kemp, M. (1969) *The Pure Theory of International Trade and Investment* (Englewood Cliffs: Prentice Hall).

Legi-Slate. 'On-line Congressional Service', (Washington, DC).

Lustig, N., B. Bosworth and R. Lawrence (eds.), (1992) *North American Free Trade* (Washington, DC: Brookings Institution).

Magee, S., W. Brock and L. Young (1989), *Black Hole Tariffs and Endogenous Tariff Policy* (New York: Cambridge University Press).

McArthur, J. and S. Marks (1988), 'Constituent Interest vs. Legislator Ideology: The Role of Political Opportunity Cost', *Economic Enquiry*, 26, 461–470.

© Blackwell Publishers Ltd. 1996

Pastor, R. (1980) *Congress and the Politics of US Foreign Economic Policy, 1929–1976* (Berkeley: University of California Press).

Passell, P. (1994), 'An End Run on GATT? Free Trade Pact is Being Watered Down as Panel Works on "Enabling Legislation"', *New York Times* (6 August), B1, B25.

Rogowski, R. (1987), 'Political Cleavage and Changing Exposure to Trade', *American Political Science Review*, **81**, 1121–1138.

Thompson, J. and G. Montcrief (1988), 'Pursuing Pork in the State Legislature', *Legislative Studies Quarterly*, **13**, 393–402.

Tower, E. and S.C. Tosini (1987), 'The Textile Bill of 1985: The Determinants of Congressional Voting Patterns', *Public Choice*, **54**, 19–26.

US Bureau of Labour Statistics (1994), *Employment and Earnings*, (March).

US Department of Commerce (1992), *US Exports to Mexico: A State-by-State Overview, 1987–1991* (Washington, DC: US Government Printing Office).

US Department of Commerce (1993a), *Statistical Abstract of the United States* (Washington, DC: US Government Printing Office).

US Department of Commerce (1992), *US Exports to Mexico: A State-by-State Overview, 1987–1991* (Washington, DC: US Government Printing Office).

Wines, M. (1993), 'From Rust Belt, Voting Yes on Trade Pact is Folly, Yes?', *New York Times* (17 November) A1, A10.

© Blackwell Publishers Ltd. 1996

40

WHARTON REPROGRAPHICS

Child Labor in Pakistan

by JONATHAN SILVERS

GILLES RIGOULET/SYGMA

Pakistan has recently passed laws greatly limiting child labor and indentured servitude—but those laws are universally ignored, and some 11 million children, aged four to fourteen, keep that country's factories operating, often working in brutal and squalid conditions

No two negotiations for the sale of a child are alike, but all are founded on the pretense that the parties involved have the best interests of the child at heart. On this sweltering morning in the Punjab village of Wasan Pura a carpet master, Sadique, is describing for a thirty-year-old brick worker named Mirza the advantages his son will enjoy as an apprentice weaver. "I've admired your boy for several months," Sadique says. "Nadeem is bright and ambitious. He will learn far more practical skills in six months at the loom than he would in six years of school. He will be taught by experienced craftsmen, and his pay will rise as his skills improve. Have no doubt, your son will be thankful for the opportunity you have given him, and the Lord will bless you for looking so well after your own."

Sadique has given this speech before. Like many manufacturers, he recruits children for his workshop almost constantly, and is particularly aggressive in courting boys aged seven to ten. "They make ideal employees," he says. "Boys at this stage of development are at the peak of their dexterity and endurance, and they're wonderfully obedient—they'd work around the clock if I asked them." But when pressed he admits, "I hire them first and foremost because they're economical. For what I'd pay one second-class adult weaver I can get three boys, sometimes four, who can produce first-class rugs in no time."

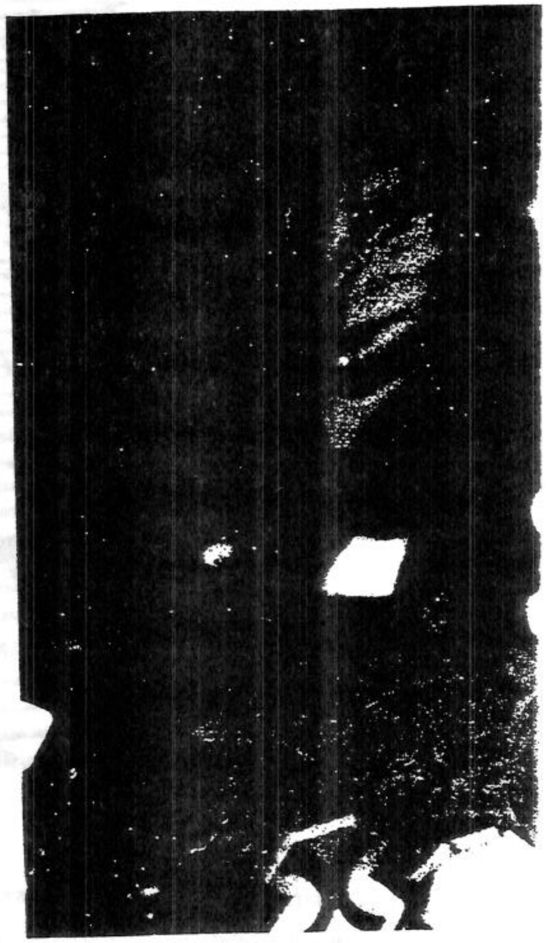
The low cost of child labor gives Sadique and his fellow manufacturers a significant advantage in the Western marketplace, where they undersell their competitors from countries prohibiting child labor, often by improbable amounts. Not surprisingly, American and European consumers are attracted to low-price, high-quality products, and imports of child-made carpets from Pakistan have trebled in the past two decades. Pakistan's carpet makers have satisfied this surging demand by expanding production at existing factories and opening new ones wherever they can. To maximize their returns, virtually all these factories employ children, and an increasing number do so exclusively. Somewhere between 500,000 and one million Pakistani children aged four to fourteen now work as full-time carpet weavers. UNICEF believes that they make up 90 percent of the carpet makers' work force.

Sadique delivers his speech at volume and accompanies it with an assortment of gestures—nods, waves, raised eyebrows—that are as theatrical as they are out of place in his shambles of a workshop. He concludes with a smile and, just in case Mirza does not appreciate his generosity, adds a wistful coda: "I wish my father had given me such an opportunity." Mirza seems doubtful, perhaps because his son is seven years old, perhaps because he has seen too many of his neighbors' children suffer through similar opportunities. But he returns Sadique's smile and says in a faint voice that he hopes Nadeem will learn enough to work one day as a journeyman weaver or, better still, to open a workshop of his own.

Whatever misgivings Mirza has at the moment are overshadowed by his poverty, which is extreme and worsening. He supports a family of five by working at a nearby kiln, molding bricks by hand for up to eighty hours a week. The work pays poorly at the best of times, and on occasion it does not pay at all. Three weeks earlier a monsoon destroyed several thousand unfired bricks that had been left drying on factory grounds. The kiln owner held the workers accountable for the damage and refused to pay them for the two weeks they had spent making the bricks. The "fine," as the owner called it, proved ruinous. Already months behind on their rent and in debt to the village merchants, Mirza and his wife concluded that the only way to avoid eviction was to bond their eldest child to one of the district's manufacturers. Sadique was their first choice: he was prosperous, his workshop was near their home, and he was rumored to have an urgent need for child laborers, which they believed would translate into a high price for Nadeem.

They were half right. The workshop has a perpetual need for children, but Sadique is unwilling to pay a premium for them. For that matter, he is unwilling to pay market rates. Having dispensed with the niceties, he offers Mirza 5,000 rupees ($146) for five years of his son's labor. It's a paltry sum—roughly two months' earnings for an adult weaver. Mirza was expecting an offer at least three times as high. "Business is off this year," Sadique says, by way of preempting Mirza's objections. "When things improve, I may be able to give you another two or three hundred. Many fathers would be glad to get half this amount."

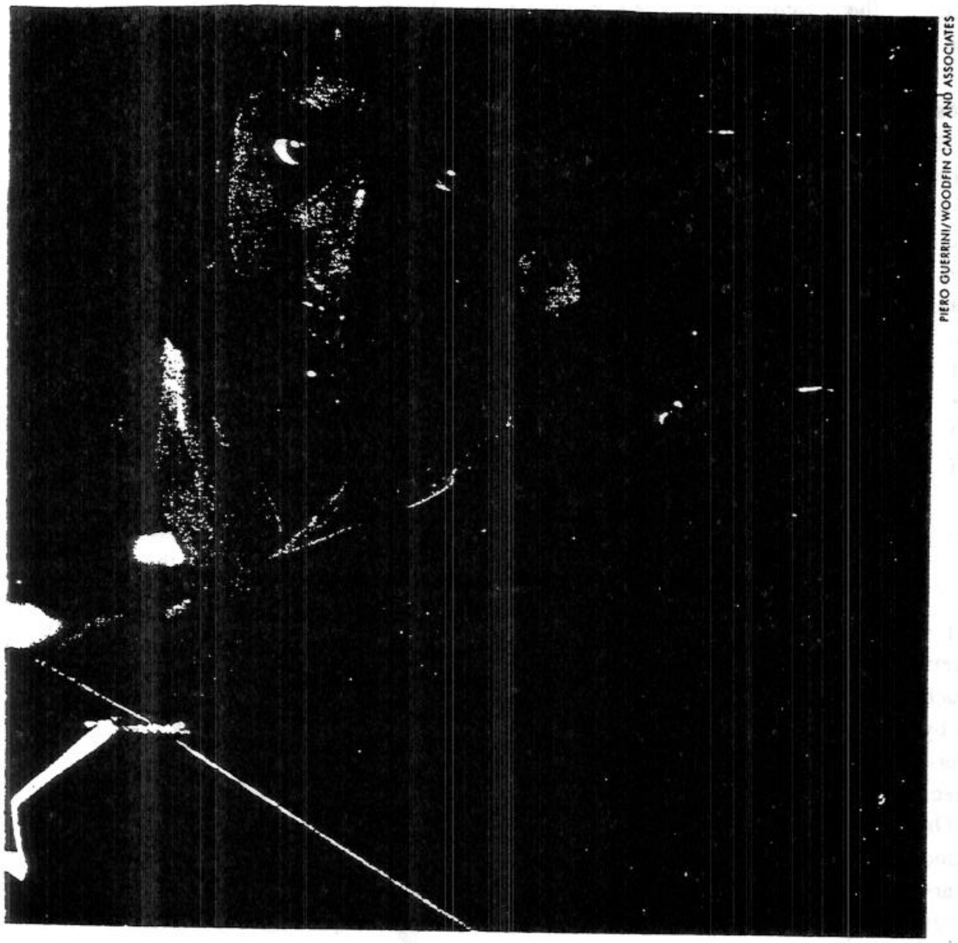
EACH YEAR MILLIONS OF CHILDREN IN PAKISTAN ENTER THE LABOR FORCE, WHERE THEY COMPETE WITH ADULTS—OFTEN EVEN WITH THEIR PARENTS—FOR WHAT LITTLE WORK IS AVAILABLE. AT LEAST HALF THESE CHILDREN ARE UNDER THE AGE OF TEN.



Previous page, a brickyard in Lahore; right, stitching soccer balls in a factory in Sialkot

Mirza is distressed. He is a small man, stooped and wasted from his years at the kiln, his skin and tunic flecked with soot. Like most laborers, he is acutely aware of his caste, and in the presence of those whom he deems his betters is deferential to the point of abjectness. Bravely he asks Sadique for another thousand rupees, though he couches the request in the most self-deprecating terms he knows. "Sir, my family's survival depends on your charity. You will always be remembered in our prayers as our savior from beggary and destitution." To his relief, Sadique agrees at once, extending a manicured hand with a speed that suggests he was prepared to pay more and got a bargain. In any event, he can afford to be generous. The money he offers Mirza, called a *peshgi*, will be paid in installments, and he will deduct from it all costs associated with Nadeem's maintenance and training. Many of the deductions are contrived and inflated. Parents are charged for their children's food and tools, the raw materials they use, the errors they make, the amount of time the master spends "educating" them. Throughout Pakistan parents consider themselves fortunate if at the end of their child's service the master has paid them one third of the *peshgi*.

Mirza is unaware of these deductions and, eager to make his escape, does not ask questions that might complicate the proceedings. He consummates the deal by shaking Sadique's hand (after wiping his own on his tunic) and accepting from him a first installment of 200 rupees. The parties are bound only by their word: no contracts are signed; no witnesses are present. "Your boy now belongs to me," Sadique says as Mirza pockets the banknotes. "Please understand that so long as he works under my roof he is answerable only to me. Inform him that the needs of my shop take priority over those of his family, and he must do all he can to please me. If he does not, we will all be disappointed, him most of all." Mirza thanks the master for his kindness, bows low, and runs off to relay this information to his son.



PIERO GUERRINI/WOODFIN CAMP AND ASSOCIATES

An Inexhaustible Labor Pool

CHILD labor has assumed epidemic proportions in Pakistan. Statistics are unreliable, but the Human Rights Commission of Pakistan (HRCP) last year estimated the number of Pakistani working children to be "realistically in the region of 11–12 million." At least half these children are under the age of ten. Despite a recent series of laws prohibiting child labor and indentured servitude, children make up a quarter of the unskilled work force, and can be found in virtually every factory, every workshop, every field. They earn on average a third of the adult wage. Certain industries, notably carpet making and brick making, cannot survive without them. One World Bank economist maintains that Pakistan's economic viability correlates with the number of children in its factories. The child labor pool is all but inexhaustible, owing in part to a birth rate that is among the world's highest and to an education system that can accommodate only about a third of the country's school-age children. Each year millions of children enter the labor force, where they compete with adults—often even with their parents—for what little work is available. In many regions the surplus of cheap child labor has depressed the already inadequate adult wage to the point where a parent and child together now earn less than the parent alone earned a year ago. As long as children are put to work, poverty will spread and standards of living will continue to decline.

To be sure, child labor is an institution throughout the Third World, and its incidence has been increasing in countries that are usually described as advanced. The worldwide population of children under fourteen who work full-time is thought to exceed 200 million. But few countries have done less to abolish or to contain the practice than Pakistan. And fewer still have a ruling class that opposes workplace reform and human-rights initiatives as vigorously. Given its relative prosperity, its constitutional prohibition against child labor, and its leaders' signatures on every UN human- and child-rights convention, Pakistan's de facto dependency on child labor is troubling and to its critics inexcusable.

"Inaction speaks louder than words," says I. A. Rehman, the director of the HRCP. "This government is in continuous violation of the Convention on the Rights of the Child, and has consistently refused to enforce those very laws it enacted to protect its most vulnerable citizens. We have far more in the way of resources and legal remedies than China, India, and Indonesia, and we do far less for our young than they. The problem is lack of political will. The problem is greed."

The median age of children now entering the Pakistani work force is seven. Two years ago it was eight. Two years from now it may be six. In the lowest castes, children become laborers almost as soon as they can walk. Much of the nation's farmland is worked by toddlers, yoked teams of three-, four-, and five-year-olds who plough, seed, and glean fields from dawn to dusk. On any given morning the canal banks and irrigation ditches in rural villages are lined with urchins who stand no taller than the piles of laundry they wash for their wealthier neighbors. Even the world-class industries of Islamabad, the modern capital, are staffed in large part by children and adolescents; politicians traveling to the National Assembly can't help noticing the ragged youths entering and exiting the brick factories, steel mills, and stone-crushing plants at all hours of the day and night. These children work with a minimum of adult supervision. An overseer comes by periodically to mark their progress and to give them instructions or a few encouraging blows, but for the better part of the workday they are left to themselves. "Children are cheaper to run than tractors and smarter than oxen," explains one Rawalpindi landowner. He prefers field hands between seven and ten years old, "because they have the most energy, although they lack discipline."

In rural areas children are raised without health care, sanitation, or education; many are as starved for affection as for food. As soon as they're old enough to have an elementary understanding of their circumstances, their parents teach them that they are expected to pay their way, to make sacrifices, and, if necessary, to travel far from home and live with strangers. "When my children were three, I told them they must be prepared to work for the good of the family," says Asma, a Sheikhupura villager who bonded her five children to masters in distant villages. "I told them again and again that they would be bonded at five. And when the time came for them to go, they were prepared and went without complaint."

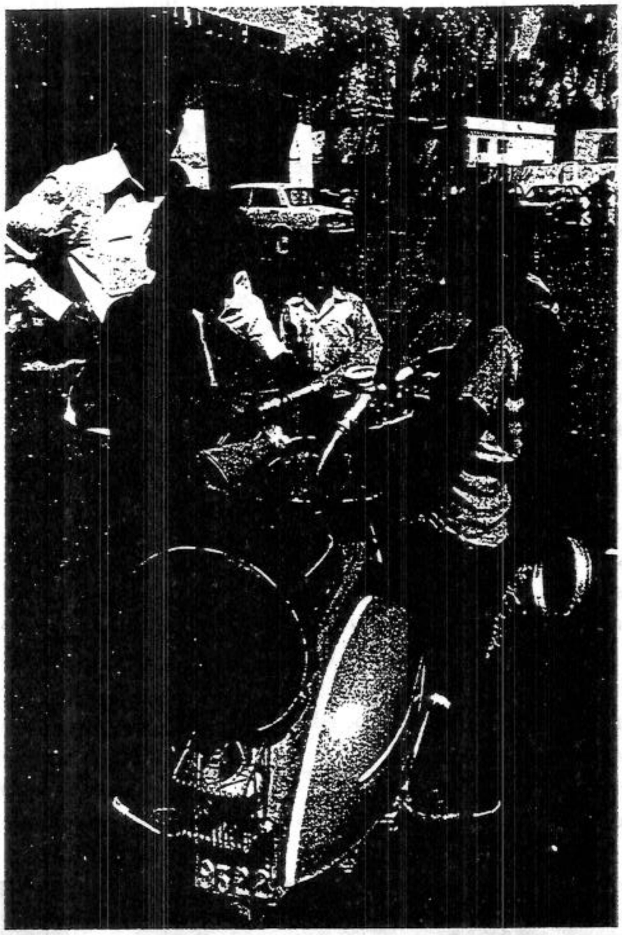
Bonding is common practice among the lower castes, and although the decision to part with their children is not made lightly, parents do not agonize over it. Neither, evidently, do the children, who regard bonding as a rite of passage, the event that transforms them into adults. Many look forward to it in the same way that American children look forward to a first communion or getting a driver's license. They are eager to cast off childhood, even if to do so means taking on adult burdens. Irfana, a twelve-year-old schoolgirl who spent four years as a brick worker before she was freed by an anti-slavery organization, remembers feeling relieved when her father handed her over at age six to a kiln owner. "My friends and I knew that sooner or later we'd be sent off to the factories or the fields. We were tired of doing chores and minding infants. We looked forward to the day when we'd be given responsibilities and the chance to earn money. At the time work seemed glamorous and children who worked seemed quite important."

She soon learned otherwise. "For the masters, bonded children are a commodity. My master bought, sold, and traded us like livestock, and sometimes he shipped us great distances. The boys were beaten frequently to make them work long hours. The girls were often violated. My best friend got ill after she was raped, and when she couldn't work, the master sold her to a friend of his in a village a thousand kilometers away. Her family was never told where she was sent, and they never saw her again."

Early in this decade the Pakistan National Assembly enacted two labor laws meant to curb such practices. The first, The Employment of Children Act of 1991, prohibited the use of child labor in hazardous occupations and environments. The second, The Bonded Labor Act of 1992, abolished indentured servitude and the *peshgi* system. As progressive as these laws were, the government failed to provide for their implementation and enforcement. It also neglected to inform the millions of working children and indentured servants that they were free and released from their debts. "We prefer to leave enforcement to the discretion of the police," says a Ministry of Labor official. "They understand best the needs of their community. Law is not an absolute. We must expect a certain flexibility on the part of those who enforce it. Could this sometimes mean looking the other way? Absolutely."

A Diminutive Entrepreneur

THE farther authorities are from a major city in Pakistan, the less likely they are to pursue violators of the child-labor laws. To leave Lahore, the nation's intellectual and commercial center, is to enter a land populated and run by children. The change is as abrupt as it is extreme. The roads just beyond the city limits are congested with donkey carts, all of them driven by teamsters of eight or nine. Boys seem to have a monopoly on roadside attractions: gas stations, auto-repair centers, restaurants. When I pull into the Star Petroleum station on



J. P. LAFFONT/SYGMA

Young attendants at a gas station

the Ferozpur Road, five miles from Lahore, three boys rush out of the garage to service my car. They are twelve, eight, and seven, and wear uniforms intended for men twice their size. The eldest has rolled up his pants and sleeves, but his colleagues helplessly trail theirs in the dirt. While the older boys fill my tank with a rusted hand pump, the youngest climbs onto the hood and cleans the windshield with a dangling sleeve. When I pull away, the boys rush back to the garage and to a diesel engine they are attempting to rebuild between fill-ups. No adults are visible on the premises.

Adults are also in short supply at the crossroads markets that provide villagers with everything from prayer mats to surgical instruments. Twelve of the fifteen stands at the Tohkar Road market are managed by children under fourteen. The fruit stand is run by a tyrannical eight-year-old boy and his four- and five-year-old sisters. The boy spends his morning slicing melons with a knife half his size, while behind him the girls sort cartloads of fruit. At the next stall two eleven-year-old cousins fashion sandals out of discarded tires. They work from dawn to dusk six days a week, and make more than 1,200 pairs each week. Behind the last stall another boy is struggling to unload a stack of carpets from his donkey cart. He weighs seventy pounds. The twenty-odd carpets in his cart weigh sixty pounds apiece, and it takes him ten minutes of yanking, hefting, and cursing to get each one into the stall. The stall's proprietor watches him with interest, but his concern is strictly for the merchandise. He is a tall, heavyset forty-year-old who looks as if he could unload the entire cart in fifteen minutes without breaking a sweat. But he makes no move to help the boy, and seems to regard his exertions as routine. So do the passersby. And, for that matter, so does the boy.

His name is Faiz. A lively nine-year-old, he has been working as a hauler since he was six. He attended school for two years, but dropped out when an elderly neighbor offered him an advantageous lease on the cart and donkey. He runs the business alone, and spends his days scrounging for hauling jobs and shuttling produce, scrap metal, and crafts around six villages. He averages sixty miles a week—no easy feat with a donkey that trots at three miles an hour. "The work is painful and the days are long, but I earn enough to feed myself and tend the donkey," Faiz says with an entrepreneur's pride. The key to his success is underbidding the competition; his rates are a tenth of his predecessor's. "It is reasonable that people should pay me less. My equipment is the same as an adult's, but I am small and have a fraction of an adult's strength. I take longer to make deliveries, so I must charge less. My hope is that the more goods I move, the stronger I will get and the more I can charge."

Soon after I arrived in Pakistan, I arranged a trip to a town whose major factories were rumored to enslave very young children. I found myself hoping during the journey there that the children I saw working in fields, on the roads, at the marketplaces, would prepare me for the worst. They did not. No amount of preparation could have lessened the shock and revulsion I felt on entering a sporting-goods factory in the town of Sialkot, seventy miles from Lahore, where scores of children, most of them aged five to ten, produce soccer balls by hand for forty rupees, or about $1.20, a day. The children work eighty hours a week in near-total darkness and total silence. According to the foreman, the darkness is both an economy and a precautionary measure: child-rights activists have difficulty taking photographs and gathering evidence of wrongdoing if the lighting is poor. The silence is to ensure product quality: "If the children speak, they are not giving their complete attention to the product and are liable to make errors." The children are permitted one thirty-minute meal break each day; they are punished if they take longer. They are also punished if they fall asleep, if their workbenches are sloppy, if they waste material or miscut a pattern, if they complain of mistreatment to their parents or speak to strangers outside the factory. A partial list of "infractions" for which they may be punished is tacked to a wall near the entrance. It's a document of dubious utility: the children are illiterate. Punishments are doled out in a storage closet at the rear of the factory. There, amid bales of wadding and leather, children are hung upside down by their knees, starved, caned, or lashed. (In the interests of economy the foreman uses a lash made from scrap soccer-ball leather.) The punishment room is a standard feature of a Pakistani factory, as common as a lunchroom at a Detroit assembly plant.

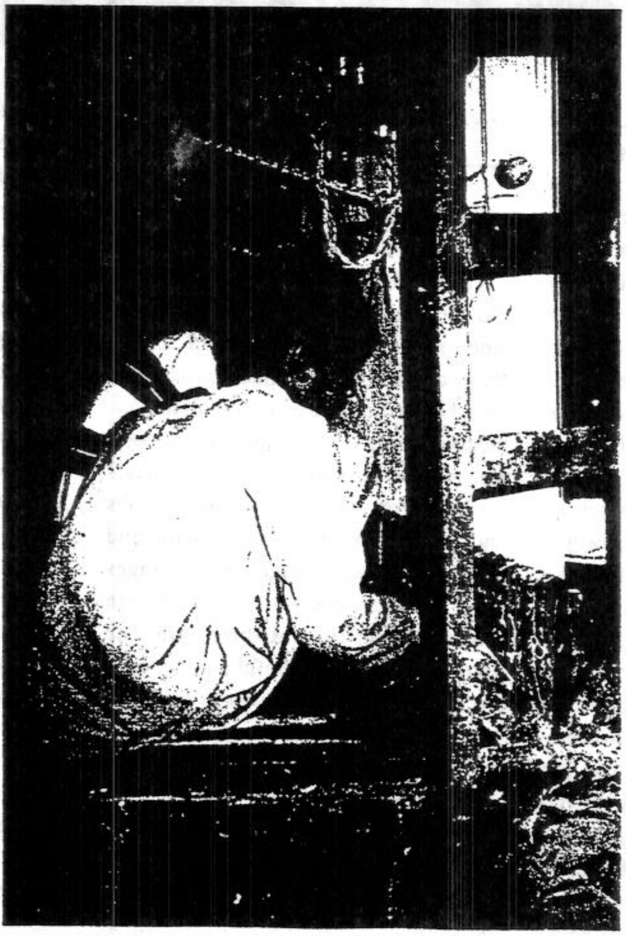
The town's other factories are no better, and many are worse. Here are brick kilns where five-year-olds work hip-deep in slurry pits, where adolescent girls stoke furnaces in 160° heat. Here are tanneries where nursing mothers mix vats of chemical dye, textile mills where eight-year-olds tend looms and breathe air thick with cotton dust.

When confronted with questions from a foreigner about their use of child labor, industrialists respond in one of two ways: they attack the questioner or they deliver a lengthy lecture about the role of children in Pakistan's development. The attacks are not always verbal. Last June a Norwegian trade-union delegation was attacked at the Sialkot sporting-goods factory by three or four armed men who were believed to work for the factory's owner. The delegation's guide and cameraman were severely beaten and the latter required hospitalization. The police characterized the attackers as "civic-minded" and warned the delegation against inspecting other area factories and "unnecessarily antagonizing factory owners."

More common, though, is the industrialist who ushers the foreign investigator into his office, plies him with coffee and cake, and tells him in his friendliest manner that child labor is a tradition the West cannot understand and must not attempt to change. "Our country has historically suffered from a labor shortage, a deficit of able-bodied men," says Imran Malik, a prominent Lahore carpet exporter and the vice-chairman of the Pakistan Carpet Manufacturers and Exporters Association. "Children have compensated for this shortage. They have worked when adults could not. They have helped construct Pakistan's infrastructure and advanced its industry. For thousands of years children have worked alongside their parents in their villages. The work they now do in factories and workshops is an extension of this tradition, and in most ways an improvement on it. The children earn more than they would elsewhere. They contribute significantly to their family's security and raise their standard of living."

The industrialist's argument is accurate only in its assertion that Pakistani children have traditionally worked with their families. But children seldom worked *outside* the family until the 1960s, when the Islamic Republic made a dramatic effort to expand its manufacturing base. This led to a spectacular and disproportionately large increase in the number of children working outside the home, outside the village, at factories and workshops whose owners sought to maximize profits by keeping down labor costs. The rise in child abuse was as meteoric as the rise in child labor. The children working in these factories were beyond the reach or care of their families and were increasingly the victims of industrial accidents, kidnapping, and mistreatment.

At the carpet looms of Wasan Pura



JONATHAN SILVERS

A Mixed Curse

"If employers would apply as much ingenuity to their manufacturing processes as they do to evading labor laws, we'd have no child-labor problem," says Najanuddin Najmi, the director general of the Workers Education Program, a government agency. "There's little doubt that inexpensive child labor has fueled Pakistan's economic growth. Entire industries have relocated to Pakistan because of the abundance of cheap child labor and our lax labor laws. At the same time, child labor has hindered our industrial development, especially in the use of advanced technologies. Why should a manufacturer invest in labor-saving technology when labor-intensive mechanisms are so much cheaper? We are discovering more and more factories that have been redesigned and retooled so that only children can work there."

Child labor has been a mixed curse for all of southern Asia, expanding its industrial capacity while generating an unprecedented assortment of social problems. Not surprisingly, Pakistan's leaders are of two minds on the subject. Speaking officially, they deplore the practice and have nothing but pity for the roughly 11 million children working in factories, in fields, and on the streets. Speaking pragmatically, they regard the practice as a distasteful but unavoidable part of an emerging economy which time and prosperity will end. They are quick to take offense (and quicker to take the offensive) when human-rights activists suggest that they have ignored the problem.

"Westerners conveniently forget their own shameful histories when they come here," says Shabbir Jamal, an adviser to the Ministry of Labor. "Europeans addressed slavery and child labor only after they became prosperous. Pakistan has only now entered an era of economic stability that will allow us to expand our horizons and address social concerns. Just as we are catching up with the West in industrial development, so we are catching up in workplace and social reforms. We are accelerating the pace of reform and have resolved to create viable welfare and educational strucures that will eradicate child labor in he foreseeable future."

Foreseeable may be a long way off. At the moment Prime Minister Benazir Bhutto seems more interested in outitting her army than in reforming Pakstani society; her government has embarked on an ambitious military buildup that has already imperiled the egion. Its first victims have been Pakstan's lower castes, the working poor who are accustomed to receiving little n the way of social services and must now make do with less. In 1994 miliary spending was 240 percent as high as spending on health and education ombined; the disparity is expected to widen in years to come. Spending on education remains among the world's owest. Only 37 percent of Pakistan's 25 million school-age children complete primary school—as compared with a world average of 79 percent and a South Asian average of approximately 50 percent. By the year 2000 less than a third of Pakistani children will attend school. The rest will enter the work force or become beggars.

Behind these statistics lurks an unpleasant truth: despite its modern views on warfare and industrialization, Pakistan remains a feudal society, committed to maintaining traditions that over the centuries have served its upper castes well. The lords—factory owners, exporters, financiers—reflexively oppose any reforms that might weaken their authority, lower their profit margins, or enfranchise the workers. "There is room for improvement in any society," the industrialist Imram Malik says. "But we feel that the present situation is acceptable the way it is. The National Assembly must not rush through reforms without first evaluating their impact on productivity and sales. Our position is that the government must avoid so-called humanitarian measures that harm our competitive advantages." On those rare occasions when a reform does squeak through, the backlash is fierce. For example, when the legislature last year approved a modest tax on bricks to fund an education program, brick-kiln owners staged a ten-day nationwide protest and threatened to suspend production, crippling construction, until the tax was repealed. Trade associations have used similar strong-arm tactics to fight minimum-wage legislation, occupational-safety regulations, and trade-union activity.

"The Charter of Freedom"

WITH a government that is at best ambivalent about social issues and an industrial sector resistant to workplace reform, the task of abolishing child labor has fallen to the human-rights community. But in a country where corruption is pervasive and education scarce, social activists are everyone's natural enemy. The ruling class despises them for assaulting its profitable traditions. The lower castes suspect them of ulterior motives. (Laborers are forever asking activists, "Why would an educated man trouble himself with the poor?") Consequently, activists are frequent targets of slander, police harassment, and lawsuits. They are beaten just as frequently, and on occasion they are killed.

Yet they persist, and sometimes they prevail. If human-rights organizations are judged by the number of people they have helped, the Bonded Labor Liberation Front is probably the most successful in Pakistan. Since its founding, in 1988, the BLLF has led the fight against bonded and child labor, liberating 30,000 adults and children—frequently entire families—from brick kilns, carpet factories, and farms, and placing 11,000 children in its own primary school system (its motto: "Struggle against slavery through education"). At the same time, it has won 25,000 high-court cases against abusive and unscrupulous employers, and helped to push the recent labor legislation through the National Assembly.

"Our victories amount to a hardship," says Ehsan Ulla Khan, the BLLF's founder and guiding force. "The state has done nothing to enforce the anti-slavery laws or even to inform the public that child and bonded labor have been outlawed. It's evident that if the enslaved workers are to be delivered from bondage, private citizens will have to do the delivering. That is, we will have to proclaim the end of slavery, educate workers, monitor employer compliance, and take legal action when necessary, because the state lacks the will and resources to do so."

With little funding, the BLLF wages a two-front war against enterprises that use child and bonded labor. While its legal advisers engage the courts and the legislature, its field staff shuttles around the country, informing workers of their recently acquired rights and distributing a pamphlet known as "The Charter of Freedom," which enumerates those rights in simple language. If a bonded laborer—child or adult—asks for its help, the BLLF takes whatever legal action is necessary to secure his or her release.

These days a surprising number of workers are refusing the pamphlet and turning their backs on BLLF staff members. This is an expression less of ingratitude than of fear. Employers throughout Pakistan are cautioning their workers against consorting with reformers who spread "false rumors" about the end of bonded labor. Many workers have been threatened with dismissal or violence if they speak with "the abolitionists" or are caught with "illegal communist propaganda."

So effective is the factory owners' disinformation campaign that workers literally flee when approached by BLLF staff members. This happened recently outside a Muridke brick factory to a BLLF leader I'll call Tariq. The fifty-odd kiln workers leaving the factory at the end of the workday scattered in all directions when they noticed Tariq lingering outside the factory gate, pamphlets in hand. One soot-covered girl of eight, left behind in the confusion, burst into tears when Tariq asked if she needed help. Between sobs the girl pleaded, "Please, sir, I have nothing to tell you. Please let me go."

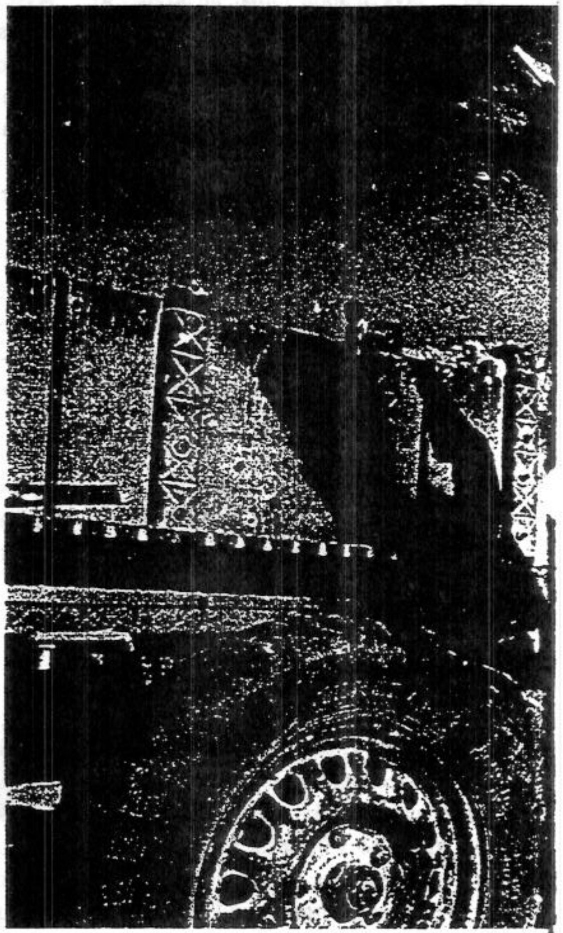
Tariq did, albeit reluctantly. He has witnessed scenes like this countless times; they happen more and more often. If they discourage him (how could they not?), he takes care not to let anyone know. He describes his work as "an outgrowth of my patriotism." "What we do is meant not to shame Pakistan before the world but to create a Pakistan that respects the rights of all its peoples and encourages human potential." Tariq is a tall, pensive thirty-nine-year-old, an artist by training and by temperament. He traces his interest in child labor to an afternoon five years ago when an anti-slavery activist entered his graphic-design studio in need of a brochure for his struggling organization. "Ehsan Ulla Khan had little money to spare, and he intimated that he'd rather not pay at all for the design work," Tariq told me. "I was just starting out in business and had no interest in politics or human rights. But I was moved by his photos of the children and agreed to do the work." Within six months Tariq was preparing all of the BLLF's documents; within a year he was overseeing its operations. Today he is its factotum: equal parts tactician, recruiter, instructor, fundraiser, morale-booster.

Some days he is also part spy. In addition to their assigned duties, the BLLF's 600 staff members are encouraged to spend their free time scrounging for leads on factory owners who are especially abusive to children. All rumors are passed on to the BLLF's Lahore headquarters. Tariq does what he can to substantiate the worst of them, usually by touring the factories. It's a duty he dislikes. For one thing, it's exhausting: there are too many leads, too many rumors to verify. For another, it's dangerous: he's had numerous clashes with publicity-shy employers and their thugs. He prefers to travel alone, reasoning that one man is less conspicuous and less of a threat than is a group. And despite his reservations he is adept at subterfuge, at gaining entry to factories by masquerading as a laborer, a wholesaler, an exporter. "I do not misrepresent myself," he says. "But if a foreman mistakes me for a businessman or a wholesaler, I don't correct him."

His first stop one day last summer was a carpet workshop in a village twenty-four miles from Lahore. The village amounted to thirty brick huts, and the workshop was small in proportion—about the size of a subway car, and about as appealing. The long, narrow room contained a dozen upright looms. On each rough-hewn workbench between the looms squatted a carpet weaver. The room was dark and airless. Such light as there was came from a single ceiling fixture, two of its four bulbs burned out. A thermometer read 105°, and the mud walls were hot to the touch. A window promised some relief, but it was closed against fabric-eating insects.

Tariq entered quietly, in slacks, shirt, and patent-leather loafers. This outfit is uncommon in the provinces; he hoped

IN THE LOWEST CASTES, CHILDREN BECOME LABORERS ALMOST AS SOON AS THEY CAN WALK. BY THE YEAR 2000 LESS THAN A THIRD OF PAKISTANI CHILDREN WILL ATTEND SCHOOL. THE REST WILL ENTER THE WORK FORCE OR BECOME BEGGARS.



Boys in a quarry do the work of machines, hauling stones for highway construction it marked him as a person with Western tastes, and his vehicle, a Toyota Land Cruiser (donated to the BLLF by UNICEF), which he had parked conspicuously close to the entrance, marked him as a man of means—a buyer, a broker, an exporter. The weavers smiled at him, and a few bowed, but no one dared speak to him. Tariq took advantage of their reverence—and the master's absence—by circling the room, noting its conditions. After two circuits he began guessing the ages of the young weavers: "Are you twelve?" The boy nodded. Tariq pointed to the next. "Fourteen?" Another nod and a smile. "Ten?" This time the nod was shy, and someone mentioned that the day before had been the boy's birthday. Tariq wished him health and happiness.

Of the twelve weavers, five were eleven to fourteen, and four were under ten. The two youngest were brothers named Akbar and Ashraf, aged eight and nine. They had been bonded to the carpet master at age five, and now worked six days a week at the shop. Their workday started at 6:00 A.M. and ended at 8:00 P.M., except, they said, when the master was behind on his quotas and forced them to work around the clock. They were small, thin, malnourished, their spines curved from lack of exercise and from squatting before the loom. Their hands were covered with calluses and scars, their fingers gnarled from repetitive work. Their breathing was labored, suggestive of tuberculosis. Collectively these ailments, which pathologists call captive-child syndrome, kill half of Pakistan's working children by age twelve.

Tariq and I watched Akbar in silence for some time. A hand-knotted carpet is made by tying short lengths of fine colored thread to a lattice of heavier white threads. The process is labor-intensive and tedious: a single four-by-six-foot carpet contains well over a million knots and takes an experienced weaver four to six months to complete. The finest, most intricate carpets have the highest density of knots. The smaller the knot, the more knots the weaver can cram into his lattice and the more valuable the finished carpet. Small knots are, of course, made most easily by small hands. Each carpet Akbar completed would retail in the United States for about $2,000—more than the boy would earn in ten years.

GILLES RIGOULET/SYGMA

Observing a child carpet weaver at work generates in an American alternating currents of admiration and anger. At one moment the boy seems a prodigy, his carpet a lesson in geometry and colors. His patience is remarkable; his artistry seems effortless and of the highest order—comparable to, say, that of a great medieval tapestry master. The next moment he fumbles with his scissors, and one notices a welt on his forearm. Suddenly the monotony of tying thousands of threads each hour seems like torture of the worst sort—like a death sentence, which in a way it is.

After ten minutes Tariq knelt by Akbar's side and said softly, "You're very good at this. The master must be quite pleased with you." The boy shook his head and grimaced. "The master says I am slow and clumsy."

Tariq placed a sympathetic hand on the boy's shoulder. "Have you been punished for poor work?" he asked. The boy shrugged and tied a red knot. Tariq repeated the question. This time the boy tied a dozen knots before answering him, in a conspiratorial whisper. "The master screams at us all the time, and sometimes he beats us," he said. "He is less severe with the younger boys. We're slapped often. Once or twice he lashed us with a cane. I was beaten ten days ago, after I made many errors of color in a carpet. He struck me with his fist quite hard on the face." By way of corroborating this, Akbar lifted a forelock, revealing a multicolored bruise on his right temple. Evidently the master did not consider the blow sufficient punishment: "I was fined one thousand rupees and made to correct the errors by working two days straight." The fine was added to Akbar's debt, and would extend his "apprenticeship" by several months.

"Do you like working here?"

"Oh, no, sir, staying here longer fills me with dread. I know I must learn a trade. But my parents are so far away, and all my friends are in school. My brother and I would like to be with our family. We'd like to play with our friends. This is not the way children should live."

Tariq listened to this outpouring without emotion. He has cultivated what he calls a surgeon's insensitivity to ravaged flesh, "because otherwise my heart would break ten times a day." Neither Akbar nor the others knew that child labor was illegal, that they were free to leave the workshop whenever they wished.

Tariq left the factory and, on a whim, headed for the district police headquarters. As a rule BLLF members are closely observant of legal procedure, lest they be accused of subversive activity. The organization's legal advisers typically spend weeks drafting a formal complaint against a factory, based on members' espionage, before they register it with a high-court magistrate. Right now, however, Tariq was as interested in testing the responsiveness of the police as in penalizing the factory owner.

The nearest police station is a colonial relic on the Lahore road in Muridke. Tariq was caught up in the usual bureaucratic chaos on entering. The foyer was packed with police officers, soldiers, crime victims, and criminals, half of them shouting, the other half covering their ears against the noise. Every now and then the soldiers tried to impose order on the crowd, but with tattered uniforms and clipless rifles their authority went only so far. Familiar with such outposts, Tariq took his place in a line and forty minutes later was face-to-face with the district sergeant. It was ten in the morning. The sergeant had been at his post for two hours, but it could have been 200 for the way he looked. Tariq told him about the conditions in the workshop, about the children. The sergeant was perplexed. "Is this a crime?" he asked. "No one has ever complained before. What do you want us to do about it?" Tariq suggested sending officers to investigate, along with a medical-services crew for the children.

The sergeant left to consult his superior. Two minutes later he returned with the superintendent, a gracious, mustachioed man of fifty. "We are not unsympathetic to your complaint," the superintendent informed Tariq. "But the place you describe is registered as a home enterprise. It is run by a small landowner, and the workers are his immediate family. Family businesses are exempt from the labor laws. This enterprise is not illegal." The superintendent opened a binder and showed Tariq the workshop's registration certificate. Tariq attempted to correct him, but the superintendent said, "What you say may or may not be true. Unfortunately, our jurisdiction does not include child labor. I have no authority to investigate a private workplace. I have no evidence that the children are working there against their will or that their lives are in jeopardy. The mechanism for doing what you ask simply does not exist here."

Tariq was not disappointed, nor was he surprised. He expected no better, and was even pleased that he had rated an audience with the superintendent. Corruption is pervasive in the justice system: for a small consideration the police will look the other way when employers misuse their workers. In several districts the police are notorious for colluding with employers—supplying factories with children who have been abducted from itinerant poor families, orphanages, schools. Not long ago a boy of nine escaped from an abusive landowner and sought help from a police sergeant at this very station. The boy claimed that he had been held captive and tortured; he begged the police to return him to his parents. Instead the sergeant ordered the "fugitive" returned in shackles to the landowner. The sergeant later made the landowner a gift of the shackles, suggesting that they be used on other disruptive children.

The Death of Iqbal Masih

In 1992 Pakistani carpet exports fell for the first time in two decades. The fall was slight in absolute terms—no more than three or four percentage points—but it indicated that Western consumers were shying away from luxury goods made by Third World children. Carpet makers' fears were confirmed when in 1993 and 1994 sales fell sharply in several of the largest markets for Pakistani exports. Since carpets were an important source of foreign currency, the decline sent shock waves throughout the Pakistani economy. At a 1993 conference, officials of the Pakistan Carpet Manufacturers and Exporters Association blamed the decline on "subversive domestic organizations which are conducting misleading and false international media campaigns abroad about the use of child labor in our manufacturing processes." The conference concluded on an optimistic note: "The memory of Western consumers is brief and our enemies' meager resources cannot sustain their destructive campaign for much longer."

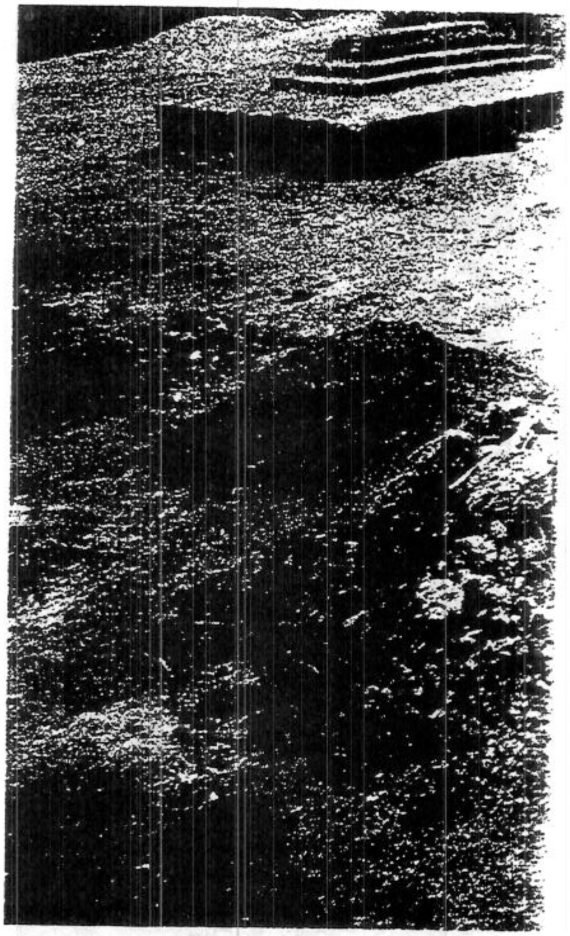
Whatever hopes the carpet makers had for a reversal of their misfortunes were dashed in 1994, when human-rights organizations around the world acclaimed a twelve-year-old former slave named Iqbal Masih for his crusade against child labor. A small, sickly boy, Iqbal had been bonded at age four to a village carpet maker. He spent much of the next six years chained to a loom, which he worked fourteen hours a day, six days a week. He was fed just enough to keep him functioning, and was beaten more often than the other children at the workshop, because, unlike them, he defied the master time and again, refusing to work and on occasion attempting to escape. At ten he slipped his chains and sought the help of the BLLF, which secured him his freedom and a place in a primary school.

Frail as he was, Iqbal was a child of rare gifts, possessed of an intellectual maturity beyond his years and a precocious sense of justice. He applied these gifts to the anti-slavery movement, and achieved results that would be impressive for a Nobel laureate, let alone a schoolboy. By his twelfth birthday he had helped to liberate 3,000 children from bondage at textile and brick factories, tanneries, steelworks—industries at the heart of the Pakistani economy. He was subsequently honored by the International Labor Organization, in Sweden; by Reebok, which presented him with its prestigious Human Rights Youth in Action Award (for "his courage and ingenuity in righting a centuries-old wrong") in Boston in December of 1994; and by ABC News, which featured him as its Person of the Week. He used his unlikely celebrity status to remind consumers that "the world's two hundred million enslaved children are your responsibility." Subsequent to his travels millions of people in the United States and Europe searched their souls and decided that they could do without products of doubtful origin from Pakistan, India, and Bangladesh.

Iqbal attained a corresponding notoriety in Pakistan, particularly among the politicians and industrialists whose feudal practices he opposed. They responded with smear campaigns and the occasional threat of violence. Iqbal dismissed these threats, telling his friends that they encouraged him to work harder. He reasoned that grown men would harm a child only as a last resort, when their own position proved vulnerable.

On the evening of April 16, 1995, Easter Sunday, Iqbal Masih was shot dead while visiting relatives in a rural village. Immediately afterward Ehsan Ulla Khan declared that the slain youth was the victim of "a mafia conspiracy." In the days that followed, Khan embellished his conspiracy theory for anyone willing to listen. "I emphatically say that the carpet mafia is responsible for this brutal killing . . . Iqbal has become a symbol of our struggle against slavery and was not afraid to expose the inhuman practices prevailing in the carpet industry. I have no doubt that the police are also a part of the conspiracy." However, Khan did not support his fulminations with evidence. "I do not rely on evidence," he told his critics. "I have my instinct. How else do you explain how, in a village where no murder has occurred for a decade, the one child who poses a threat to the carpet owners is gunned down? Coincidence is never so cruel." To the claim of the local police that Iqbal's murder was an isolated incident Khan retorts, "The evidence can be found if the police could be bothered to look." The killing remains unsolved.

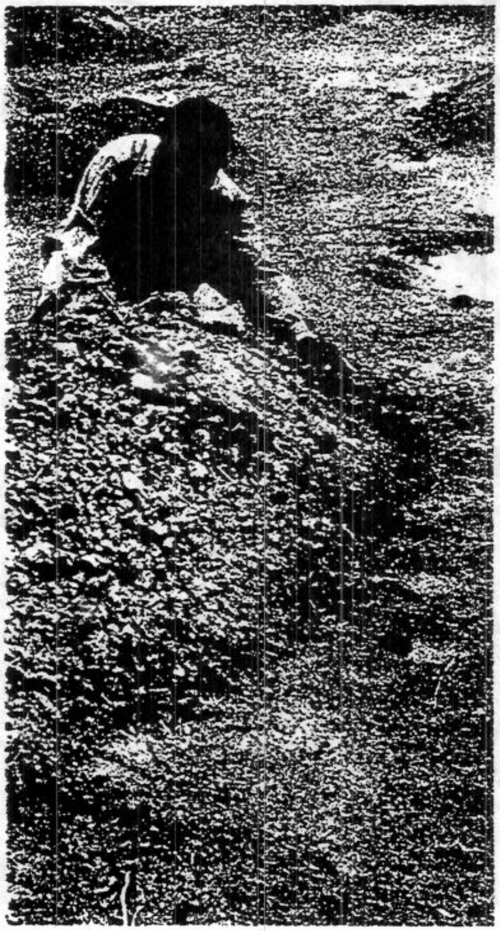
IQBAL ATTAINED NOTORIETY IN PAKISTAN, PARTICULARLY AMONG THE POLITICIANS AND INDUSTRIALISTS WHOSE FEUDAL PRACTICES HE OPPOSED. HE DISMISSED THREATS AGAINST HIM, TELLING HIS FRIENDS THAT THEY ENCOURAGED HIM TO WORK HARDER.



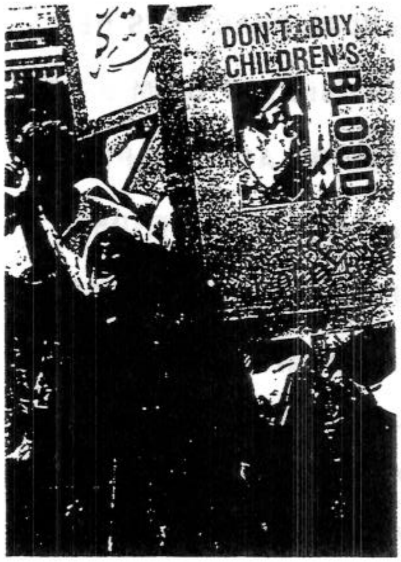
Iqbal's grave: his mother. Inayat Bibi: a protest rally after Iqbal's funeral

Eight hundred mourners crowded into the Muridke cemetery for Iqbal's funeral. A week later 3,000 protesters, half of them under twelve, marched through the streets of Lahore demanding an end to child labor. A few days after the funeral Khan left Pakistan to consult with children's-rights activists in Europe. There he repeated his accusations to great effect at conferences, on television, before lawmakers. Iqbal was proclaimed a "martyr for the cause of bonded labor"; his murder became a cause célèbre among the intelligentsia. Khan called upon the Human Rights Commission of the United Nations to ban the import and sale of all products made by children, especially carpets. "I appeal to importers and consumers: say no and only no to child-made carpets," he said. "This is the last message of Iqbal. It would be an insult to his blood and memory if people continue to buy child-made products in any part of the world."

Western consumers have responded to Khan's plea. Sales of imported carpets have fallen precipitously in recent months. Bowing to public pressure, importers in the United States, Sweden, Italy, Britain, France, and Germany by last June had canceled carpet orders collectively valued at $10 million. At the same time, human-rights groups and individual sympathizers have donated large sums to support and expand BLLF operations. Ironically, Iqbal's death opened doors and purses hat were previously closed to Khan.


JONATHAN SILVERS

JONATHAN SILVERS



ILYAS DEAN/SYGMA

Westerners, who have seen economic weapons used to achieve social reforms, might expect canceled orders to result in negotiation and, with luck, accommodation between industrialists and activists. Pakistan's industrialists, however, have chosen the questionable tactic of denying the existence of bonded labor in their factories. Shahid Rashid Butt, the president of the Islamabad Carpet Exporters Association, told his colleagues, "Our industry is the victim of enemy agents who spread lies and fictions around the world that bonded labor and child labor are utilized in the production of hand-knotted carpets. They are not and have never been." He condemned the BLLF and its allies as Jewish and Indian enemies who had launched a systematic campaign to damage the reputation of Pakistan's carpet industry for their own profit. His remarks were enthusiastically endorsed by the Pakistan Carpet Manufacturers and Exporters Association and echoed in the National Assembly.

"These charges flew in the face not just of reason but also of an extraordinary amount of evidence," says I. A. Rehman, the director of the Human Rights Commission of Pakistan. "Anywhere else they would have been laughed at and dismissed. Here they were accepted as fact and acted on." At the urging of politicians and industrialists, Javed Mahmood, the assistant director of Pakistan's Federal Investigation Agency (FIA), last May launched an inquiry into the BLLF on the strength of information he had received from highly placed sources suggesting that the organization was supported by "Pakistan's enemies." He later said, "I consider the information credible and will do all I can to protect our country's commercial interests from unscrupulous enemies." At the same time, Pakistan's leading newspapers began running "exposés" of abolitionist leaders, the nicest of which characterized Ehsan Ulla Khan as a philandering bigamist with "indisputable ties to Jewish and Indian agencies hostile to Pakistan." The publishers of these newspapers are suspected of having large financial interests in industries employing child labor.

The FIA is a secret police force, and one of its best-kept secrets is whom it works for. Nominally an organ of the state, it is not above accepting freelance assignments from prominent individuals and commercial groups. The extent of its extralegal activities is anyone's guess, but a highly respected human-rights investigator believes that "there is close cooperation between carpet interests, feudal lords, segments of the police force, and the administration—district commissioners, the courts, and government officials. Financially resourceful drug barons are also a part of the scene." Whoever the client, the FIA provides an assortment of services straight out of the KGB handbook: wiretaps, tails, searches, arrests, harassment, and varying degrees of corporal punishment.

These services were very much in evidence on a Thursday afternoon in late June, when the FIA raided the BLLF's Lahore headquarters. The detail consisted of ten men, all in plain clothes, who scrambled up four flights of stairs to the tiny office in no time flat. These were not ordinary policemen; this was not the usual surprise "inspection" (read "intimidation") to which all nongovernment organizations are periodically subjected. These were professional agents, lithe and expert, commanded by a severe officer in a freshly pressed safari suit. After lining the BLLF workers up against a wall, he ordered his troops to "confiscate anything that may incriminate them." The agents took a liberal view of "incriminate," and packed up computers, filing cabinets, fax machines, photocopiers, telephones, stationery, posters, bicycles—and the cashbox containing the monthly payroll. Their depredations were supervised by a small man who was distinctly not a policeman. He represented, it turned out, the Pakistan Carpet Manufacturers and Exporters Association. His purpose, he said, was "to protect the interests of legitimate businessmen." Every so often he consulted with the commander.

When one BLLF worker tried to protest, an agent threw her against a wall and held a rifle butt inches from her face. When another worker demanded to see a search warrant, the commander informed her that none was necessary, because "we are acting to prevent terrorism." The association representative nodded in agreement.

Fifteen minutes later the detail was gone, along with the office equipment and furnishings. All that remained was a heap of broken furniture, a workers'-rights poster, and a BLLF flag dangling out an open window. Several staff workers had been taken away as well, to an FIA holding center, where they were interrogated for three days.

Two days later another FIA detail raided the BLLF's "Freedom Campus" training facility in Lahore, along with several of its primary schools around the country. Once again the agents were undiscriminating. They seized everything movable ("items used to obstruct valid commercial interests") and mistreated the staff without respect for position or age. Teachers, drivers, secretaries, and peasant families seeking refuge from violent employers were interrogated along with administrators, advocates, attorneys, and fundraisers.

After an earlier raid on BLLF headquarters Fatima Ghulam, the director of the BLLF's women's-education program, was held for two days. "An officer promised to release me immediately if I agreed to inform against Ehsan Ulla Khan and some of the others," Ghulam says. "He wanted me to testify that Khan is a subversive, an enemy agent, and that the BLLF receives money from foreign governments. He said he had tapped my telephone conversations and had recordings of me discussing treasonable acts. If I wanted to avoid prosecution, I would have to cooperate with the FIA. I refused, and he kept me without food or water. When I wouldn't speak to him the next day, he slapped me and dragged me around the room."

Not to be outdone, the Pakistani press stepped up its campaign against the BLLF. Last summer a number of newspapers whose editorial pages conceded that they were "troubled by the carpet export crisis" reported the following "facts": Khan himself had murdered Iqbal Masih to win sympathy for the BLLF; Khan had misappropriated BLLF funds to support his own decadent lifestyle; Khan routinely used BLLF schoolchildren as sex partners and house slaves; Iqbal Masih was a twenty-one-year-old midget whom Khan paid to masquerade as a carpet child; the BLLF was an outpost of India's intelligence agency; Khan was an Indian agent working to disgrace the Pakistani carpet trade. These same papers also "revealed" that carpet workers enjoy a higher standard of living than the average citizen, along with better working conditions. "The few children working on carpets," one editorial assured its readers, "do so after school, in their own homes, under the supervision of loving parents."

In the wake of these attacks BLLF operations—child-welfare programs, schools, training and education programs—nearly shut down for lack of funds and staff. Membership has suffered, and many of the legal advisers and support staff, fearing reprisals, have fallen away. Those who remain are subject to almost constant harassment: the fortunate ones have their telephones tapped; the less fortunate are shadowed around the clock. At the same time, the courts have ignored their complaints about child labor and abusive treatment by employers.

Just in case the intention of the Federal Investigation Agency was unclear, Assistant Director Mahmood in early June charged Ehsan Ulla Khan, who was still abroad, and a BLLF strategist named Zafaryab Ahmad with sedition and economic treason, capital offenses punishable by death. According to Mahmood, "The accused men conspired with the Indian espionage agency to exploit the murder of Iqbal Masih . . . causing a recurring huge financial loss to Pakistan's business interests abroad and paving the way for India to wage economic warfare against Pakistan." Ahmad was arrested and taken to a Lahore jail, where, after repudiating the charges (he called them "foolish and absurd"), he was denied bail. The FIA has since refused to provide BLLF attorneys with evidence supporting the charges, although Mahmood assures them that it consists of "videotapes and recordings of telephone conversations that amount to firm proof." Mahmood has vowed to arrest Khan "the very moment he returns to Pakistan, the moment his aircraft touches down."

Ehsan Ulla Khan remains in Europe, an unhappy exile. "They will jail me if I return to Pakistan," he told me shortly after he left his country. "Our attorneys tell me I am of greater use to the BLLF here, speaking out against the authorities, than I would be inside a Lahore cell. I fear for my people. The police have harassed many of them, and so many more have left us out of fear. We are demoralized. We cannot pay our bills and our staff. Our schools may close and our thousands of students may end up in the very factories we saved them from. Our offices and homes are under surveillance. Our telephones are tapped. We are fighting for our survival. If the attacks do not stop soon, it is possible that the BLLF will perish. That would be tragic. What will become of the children of Pakistan?"

WHARTON REPROGRAPHICS

Public Choice 87: 379–393, 1996.
© 1996 *Kluwer Academic Publishers. Printed in the Netherlands.*

DOCUMENT DELIVERED

Regional free trade and social integration

CARSTEN HEFEKER*
Centre for European Policy Studies (CEPS), Place du Congrès 1, B-1000 Brussels, Belgium

Accepted 13 October 1995

Abstract. This paper analyzes recent free trade arrangements from a positive political economy perspective. In contrast to most other literature which fails to take into account geographical factors it is argued here that proximity and transportation costs play an important role in trade arrangements. Another important also largely neglected factor is the degree of social cohesion in terms of labor standards among potential trading partners. A political support maximizing government will consider these factors when balancing political support from free traders and protectionists.

1. Introduction

According to traditional trade theory free trade is optimal because it is welfare increasing. In reality, however, free trade is only rarely observed. Political economy theory has been able to explain why this apparent contradiction occurs. According to this approach, economic policy is made by rational and optimizing individuals who, contrary to what the traditional theory assumes, do not maximize some ad-hoc social-welfare function but their own objective function which is only in part compatible with social welfare. Policymakers try to maximize their own well-being in terms of prestige, job security and income. To this end they are influenced by special interests who offer them political support, information and sometimes even corrupt them and their political position (for a recent survey see Rodrik, 1994).

Therefore, trade policy should be seen as benefiting some important special interests and restricted trade obviously increases these groups' income in that it generates rents accruing to them. Consumers who suffer from this, because they have to pay higher prices for products or are restricted in their choice, however, do not realize nor oppose redistribution of consumer rents towards producers (disregarding the efficiency losses from protection) because these

---

* This paper was prepared while I was visiting the Lowe Institute of Political Economy at Claremont McKenna College whose hospitality is gratefully acknowledged, as is financial support from the Deutsche Forschungsgemeinschaft (SFB 178). I would like to thank Rolf Bommer and a referee for helpful comments.

losses are spread over a wide variety of goods. They are almost invisible in the household's budget and give hence no incentive to exert political pressure favoring a different trade policy. Information costs and collective action problems, which undermine the capacity to organize a large group of individuals as consumers are, contribute to the fact that there is only little pressure for free trade (see Magee et al., 1989).

In this interpretation, exporting interests are often disregarded as well, because it is assumed that they suffer from the collective action problem too. Moreover, unless some form of foreign retaliation is assumed, there is no stake for the exporting industry involved with domestic protection.[1] The exporting industry in this scenario should only lobby the foreign government for lowering foreign trade barriers (see Hillman and Ursprung, 1988). In Magee et al. (1989), however, exporters lobby their own government for export subsidies.

In recent years a shift in trade policy and theory has occurred. Now increasingly regional free trade arrangements can be observed, most notably the "Common Market Program" in Europe and more recently the North-American Free Trade Agreement (NAFTA) among the U.S., Canada and Mexico. Regional free trade has challenged the traditional theory because according to this theory universal free trade is the best and when discussing customs unions or free trade arrangements nevertheless, trade creating and trade diverting effects should be compared. The regional arrangements which can be observed, however, do not seem to follow the theoretical determination of beneficial free trade areas. Absent in the traditional theory, moreover, is the geographical factor which is clearly important in the real world.

While the regional factor is obvious, and already partly explained in the literature by gravitational factors (see e.g. Frankel and Wei, 1995) another effect became important in the negotiations between Canada and the U. S. on the one side and Mexico on the other side. One major concern voiced by interest groups was their fear of "social dumping" from the side of Mexico. By this they meant that lower social standards for the labor force in terms of working hours, social insurance and working conditions in Mexico might give Mexican suppliers an "unfair" competitive advantage vis-à-vis the American producers. Therefore, especially labor unions were afraid that NAFTA would export jobs to Mexico as American producers would shift their production South of the border to take advantage of the cheaper labor offered there. Consequently, minimum labor standards in Mexico comparable to those in the U.S. were required by American unions.

That discussion is similar to the same dispute in the European Union. Western European labor unions and their governments express the same fear with regard to South European labor conditions of the newly entering countries.

555
379

Lower labor standards in Greece, less social security to be paid by employers and child-work in Portugal are frequently cited examples. These considerations become even more important when an incorporation of more countries into the European Union is discussed. The emerging market economies in Eastern Europe or, for instance Turkey, are rejected for being seen as implying an export of jobs to these countries when lowering trade barriers towards them. In this argumentation protection is necessary to shelter industry and labor from unfair labor advantages and "social dumping".

A final example is the rapid increase of wages in the former Eastern Germany after German unification. Western unions, which conducted the wage negotiations for their new Eastern colleagues were very concerned with increasing the Eastern wages and social benefits to a degree not too much apart from the Western niveau. While the alleged reasons were "social fairness" and to avoid Eastern Germans from migrating to West Germany, it is obvious that the real reason was to avoid being confronted with a cheaper labor force in the East. Hence, wage demands in the East were also a protectionist measure for Western employees (see also Sinn and Sinn, 1993).

These two important and largely neglected issues of geography and "social cohesion" in terms of labor conditions are taken up in this paper and combined to help understand the discussion and conditions making free trade possible and viable. It is shown here that geographical proximity as well as social cohesion are influential factors in the political choice of free trade arrangements. This helps to understand the movement towards free trade from a political economy perspective because it is argued that the traditional instruments of protection, tariffs and non-tariff barriers, are substituted by another instrument, namely social integration, to reduce competitive advantages of other countries.

2. Trade policy, reciprocity and social integration

In the traditional trade policy discussion the role for protection is rooted in the exploitation of terms-of-trade effects for a large country (see Corden, 1984, for a survey). In reality, however, trade policy is used by developing and other small countries as well as by large traders who might indeed be able to exploit these effects. Notably sectors like agriculture, steel and textile are protected in most developed economies. These sectors are offered special protection because of their status as declining, i.e. shrinking in output. Protection provided to declining industries is generally explained as being founded in the judgement that specific factors employed in these declining industries, which do not have the opportunities for adjustment available to mobile factors, ought to be cushioned against income losses due to falls in their industries' output. Adjustment costs imply that factors are reluctant to move and since moral hazard causes private insurance markets for income maintenance to fail, governments are obliged to provide social insurance against income losses (Eaton and Grossman, 1985).

Hillman (1982) in contrast focuses on the role of self-interested politicians, relaxing the assumption of social welfare insurance considerations. The government in its attempt to maximize political support finds itself forced to grant protection to an industry affected by an external terms of trade shock by weighing the marginal costs of political support lost from increasing prices for one sector with the marginal support gained by granting this price increase to the negatively affected industry. For two reasons a declining industry is more likely to be protected than a booming industry. The first reason is that less firms are likely to enter a declining sector. Therefore, the gains from protection are less dispersed and the incentives to lobby for protection are higher. The second reason is that factor owners in declining industries find it more difficult to diversify into other industries. Their stakes are higher in the given industry which increases their incentive to exert political pressure. Mc Keown (1986) generalizes the case and points out that tariffs are likely to rise in a general depression when the number of declining industries increases. The tariff cycle he identifies is thus connected with the general business cycle.[2]

Given the politically beneficial use of at least temporary protection, the question arises why a self-interested government should surrender this source of gaining political support. To answer that question Grossman and Helpman (1993) analyze the viability of free trade arrangements from a political economy point of view. Politicians are perceived as maximizing an objective function which contains lobbying outlays from the industries and social welfare which is increasing under free trade. They show that the exemption of certain industries from the free trade arrangement might increase the feasibility of the arrangement. In their analysis the role of the exporting industry becomes especially important. The reason for the exporting industry to lobby its *own* government is given from the fear of retaliation from the foreign government.

Moser (1990) and Hillman and Moser (1995) focus explicitly on the role of reciprocity in striking free trade agreements. Trade liberalization is seen as guaranteeing mutual market access and is thus an exchange of rents accruing to the export industries among the involved countries. The discussion is set in a property rights concept of markets where domestic firms have a preeminent right to sell in their own markets, which is attacked by the sales of foreigners. The defence of these rights through protectionism, however, disadvantages the exporting industries and their employees, who, via reciprocity, suffer from reduced sales abroad. The export sector will therefore exert political pressure for free trade. With free trade the formerly gained tariff incomes are transformed into rents for the exporting sectors in both countries. The amount of market liberalization is then dependent on the relative political importance of the opposing industries in a country.

The fear of retaliation is also behind Schuknecht's (1992) explanation of current European integration. In his interpretation increased intra-industry trade made European countries more vulnerable to retaliatory trade policy and hence made them more interested in committed free trade. Interdependence among countries reinforces support for reducing remaining trade barriers since more companies/interest groups would profit from the reciprocal gains. Therefore intra-industrial trade should increase the relative weight of supporters of trade integration because direct reciprocity deters protection.[3]

A dynamic interpretation about the extension of free trade arrangements is given by Baldwin (1993) who explains why outside countries might want to join an existing bloc of free trading countries. Exporting industries outside the arrangements see their competitiveness and profit opportunities eroded when trade blocs are relatively closed towards external trade. This would happen when trade diversion is given or when trade blocs become more protectionist vis-à-vis outsiders. Then export industries outside the bloc begin to lobby their own governments for joining the bloc as well. Entering the arrangements increases the bloc which in turn makes other countries realize that their exporting opportunities are reduced. This dynamic effect, called domino effect by Baldwin, might help to explain why the European Community steadily increased from 6 to 9 to 12 to 15 countries.[4] It does, however, not explain why countries in the bloc might others want to join and let them into the bloc.[5]

Regionality and reciprocity are therefore usually seen as triggering a movement towards free trade and integration. As socalled gravity models show (see e.g. Bergstrand, 1985), the trade between two countries is proportionate to the product of their GNPs and inversely related to the distance between them (measured by the distance between the two capitals).[6] Since geographical proximity expands trade, this high level of trade gives rise to free trade interests because retaliation makes protection politically costly. If countries adopt harsh retaliatory measures to counter protection of trading partners, the value of such protectionist measures decreases. In contrast, by raising one's own barriers, own export chances are reduced. Thus, the balance shifts towards free trade because the costs and benefits of protection and free trade change. By moving towards free trade, however, governments lose an important instrument for gaining political support from import-competing industries. Therefore it is obvious that the viability of a free trade agreement might be increased if other instruments can substitute for tariffs to protect negatively affected industries.

One such instrument can be social integration. As already mentioned in the introduction, lobbies require social cohesion to counter the alleged "social dumping" in the cases of NAFTA and the European Union (EU) when relatively poor countries are admitted to a free trade arrangement. An obvious example is the social chapter and the protocol of social policy (annexed to the Maastricht treaty) of the EU. In this regulation an approximation of social regulation across all member countries of the EU is required. This has not happened to protect workers in countries with lower standards, but in fact aimed to "level the playing field" for international trade (see Vaubel, 1995). This social integration and common regulation ignored all differences in preferences and productivity of workers because workers' demand for leisure, hygiene, security etc. depends on income and local customs. Obviously both differ widely among European countries.

It is no coincidence that the social chapter followed the Single European Act which allowed increased mobility of goods and factors of production across European borders and established the Common Market. Politicians have thereby compensated important interest groups for market integration.[7] Increased competition has been neutralized by "raising rival's costs" (Salop and Scheffman, 1983). Interestingly enough, this aim to avoid the alleged "social dumping" is shared by unions and employers in the more developed countries.[8] Factor owners in a given sector unite in their political position and conflicts arise between sectors and not factors of production. For instance, German unions and the employers association welcomed the social charter in a joined declaration. The reason is obviously that both factors of production are industry specific and would have to bear adjustment costs when their industry suffers.

The question naturally is why less developed countries agree to such a restriction of their competitiveness. One explanation would be that they are compensated through other means as for example regional development funds in the EU case (see Lange, 1993).[9] Accepting social integration might also be a condition for admitting those countries to the agreement.

3. The model

In this section a simple model is presented which takes explicitly the influence of geography and social cohesion into account. It is a partial equilibrium model where only the interests involved with two tradable goods producing sectors are considered. The general equilibrium conflicts from trade measures, as analyzed in e.g. Magee et al. (1989), are disregarded here. Only one country is modeled where the government balances the interests of the two industries and the interests of the consumers. It is assumed that an exporting and an import-competing industry exist. The exporting industry for simplicity is assumed to sell only abroad, while the import competing industry produces for the domestic market and is under pressure from foreign competition. Therefore that industry would like to see tariffs levied on imports. The prices of the goods are determined through production and transportation costs and tariffs imposed by the domestic and the foreign government. It is moreover assumed that the supply of products of the import competing industry to the home market entails no transportation cost while the supply to the foreign market is subject to transportation costs. The transportation costs are increasing in the geographical distance between producing and consuming country.

The price of imported goods will hence reflect both transportation costs and tariffs as well as the costs of production. These in turn are given by the wage rate, w, which might differ between the two industries, indexed 1 and 2. Good 1 is produced in the exporting industry for exports alone and good 2 is import-competing. Pure production costs for both goods are similar in the domestic and the foreign country (denoted by an asterisk) and reflects the assumption that both countries operate under the same technology. The international difference in goods prices arises from different social standards and labor regulations. These are captured by the term s which in both countries applies to both domestic sectors similarly. Social regulation is thus national and not goods-specific. The price of good 1 supplied to the foreign country is hence

$$p_1 = w_1 + s + \tau + t^* \tag{1}$$

and

$$p_1^* = w_1^* + s^* \tag{2}$$

respectively. The price of good 1 is dependent on the production costs in the home country w_1, its social standards s, the transportation costs and the foreign rate of tariff t^*, where the model is formulated in logs. The competing good 1 in the foreign country is dependent on the same production costs and the social standard in that country s^*. The price of the import competing good 2 is given by

$$p_2 = w_2 + s \tag{3}$$

for domestically produced goods and

$$p_2^* = w_2^* + s^* + \tau + t \tag{4}$$

for the imported good.

Output in the two domestic industries is given as

$$y_x = \gamma(p_1^* - p_1) \tag{5}$$

and

$$y_M = \gamma(p_2^* - p_2) - \varepsilon. \tag{6}$$

The change of the output of the exporting industry, denoted by y_x, is dependent on the responsiveness of foreign demand γ for the industry's good with respect to a change in the relative price of the foreign and the domestic good. The same is true for the output of the import-competing industry, y_M. This industry, however, might also be subject to an external demand shock which affects output in this sector negatively, ε, giving an incentive to protect that sector politically as discussed in Section 2. For simplicity, only one type of shock is considered but the analysis can easily be extended to the export sector as well. A negative shock for the exporting industry would require an export subsidy for that industry.

The politically self-interested government's objective function, which can be thought of as a political support function dependent on trade policy, is given by

$$V = -\{\alpha y_x + y_M\}^2 - \mu t^2. \tag{7}$$

The government is assumed to maximize the probability of reelection subject to the political weights placed on the support of gainers and losers from a specific policy. In contrast to interest groups models (see e.g. Hillman and Ursprung, 1988; Magee et al., 1989), politicians maximize political support by balancing marginal gains and losses from a policy. Interests groups interaction is not modeled. This approach is very helpful in explaining government's behavior in response to a shift in exogenous parameters (see Hillman, 1982). Moreover, it provides the opportunity to explicitly take into account the interest of the consumers which are disregarded in lobbying approaches.

The government is interested in the trade-sectors' political support to be re-elected. The model captures this support only. Hence other political decisions are not considered as are not the interests of other sectors in general trade policy. As shown in Section 2 for the German example, interests of employers and employees are the same in this short-run model because adjustment for both factors takes time and is costly.

The first term captures the political support the government receives from the two sectors which are exposed to international trade for pursuing a policy which is beneficial to their output increases. The y_i terms are deviations of output from full employment of all factors of production and α measures the political weight of the export sector relative to the import sector. The second term captures the aversion of the general public towards tariffs levied on the imported goods. As argued, this formulation allows to include the public's interest in trade policy. The weight of price increase aversion of the public is reflected by μ. Both terms are squared to express growing political aversion with increasing deviations from the preferred output level and with tariffs. Hence, political support would be maximal with full employment and no tariffs.

3.1. *The case of non-retaliation*

First, I analyze the government's behavior in a situation with non-retaliation to contrast it in the next section with the situation of retaliation. Given a situation of initial underemployment of resources, the government is eager to expand the output of the trading sector. It can do so by increasing the rate of tariff it sets on imported goods. The export-sector is subjected to a exogenous foreign tariff, t^*, and might try to influence the tariff policy of the foreign government (as in Hillman and Ursprung, 1988) but has no stake in the domestic trade policy. Abstracting from general equilibrium effects, the export industry would be unaffected if the domestic government would raise the price of imported products since the foreign government is expected not to retaliate. This could be imagined to be the case for an extremely small open economy and should only serve as a benchmark case to be compared with the case of retaliation.

Given this situation the political-support function for the home government, by using equations (1) through (7), has the form of

$$V_N = -\{\alpha\gamma(\Delta s - \tau - t^*) + \gamma(\Delta s + \tau + t) - \varepsilon\}^2 - \mu t^2 \qquad (8)$$

where $\Delta s \equiv s^* - s$.

Notice that the transportation costs τ enter for both industry with the opposite sign, whereas the difference of social standards between the home and the foreign country Δs enters for both industries similarly. Hence both industries have different preferences regarding the geographical distance of the choice of the trading partner but unite in their wish to at least equalize social standards s and s*.

Maximizing political support for the government yields as the optimal rate of tariff

$$t_N = \frac{\gamma^2\left[\tau(\alpha-1) - \Delta s(1+\alpha) + \alpha t^*\right] + \gamma\varepsilon}{\eta} \tag{9}$$

where $\eta \equiv \gamma^2 + \mu$.

The optimal rate of tariff set by the government for the case of non-retaliation is a function of transportation costs, whereas the sign is not unambiguous. The difference in social standards in both countries, however, enters with an unambiguous sign, reflecting the common aim of the two sectors. Whenever Δs is positive the incentive to set a tariff is reduced. Both sectors' international competitiveness rises when s*>s. The larger the external shock, the larger the incentive for the government to protect the negatively affected sector. Moreover, the government increases t when t* is high.

Using (9) in (8), political support for the case of non-retaliation is given as

$$\begin{aligned} V_N = &-\left\{\Delta s\left[\gamma(1+\alpha) - \gamma^3(1+\alpha)/\eta\right] + \tau\left[\gamma(1-\alpha) - \gamma^3(1-\alpha)/\eta\right]\right. \\ &\left.-t^*[\alpha\gamma - \alpha\gamma^3/\eta] - \varepsilon[1-\gamma^2/\eta]\right\}^2 \\ &-\mu\left\{\frac{\tau\gamma^2(\alpha-1) - \Delta s\gamma^2(1+\alpha) + \gamma\varepsilon + \gamma^2\alpha t^*}{\eta}\right\}^2. \end{aligned} \tag{10}$$

3.2. *The case of retaliation*

Now assume that the structure of international trade policy changes. As explained in Section 2, the home country might no longer be in a situation in which t has no effect on the trade policy of other countries. Of course, this case is much more realistic as the danger of trade wars and frequent trade conflicts, for instance, between the EU and the U.S., or the U.S. and Japan show. Notice also that the GATT agreement explicitly foresees the measure of retaliatory protection as a way to force other countries to stick to the "rules of the game". In recent years, this retaliation has become even more important as the discussion of aggressive bilateralism shows.

Introducing retaliation changes the political trade-off for the home government because the exporting industry is now affected by *domestic* trade policy. Given a certain degree of foreign retaliation on domestic protection, $0 < \theta \leq 1$, the export industry will oppose increases in the domestic rate of tariff because it will trigger a similar response by its trading partners.

The political-support function then has the form of

$$V_R = -\left\{\alpha\gamma(\Delta s - \tau - \theta t) + \gamma(\Delta s + \tau + t) - \varepsilon\right\}^2 - \mu t^2. \tag{11}$$

This implies a politically optimal tariff rate of

$$t_R = \frac{\tau\gamma(\alpha-1)\xi - \Delta s\gamma(1+\alpha)\xi + \xi\varepsilon}{\Psi} \tag{12}$$

where $\Psi \equiv \xi^2 + \mu$ and $\xi \equiv \gamma(1-\alpha\theta)$.

Again the decisive factors are the geographical distance to the trading partner, the social standards in both countries and the size of the external shock. Given retaliation, the optimal rate of tariff is this time also influenced by the reaction of the foreign government. The higher the degree of retaliation θ, the lower the domestic tariff rate will be as reflected in ξ.

Using (12) in (11), the political support in the case of retaliation is given as

$$\begin{aligned} V_R = & -\left\{\Delta s\gamma(1+\alpha)[1-\xi^2/\Psi] + \tau\gamma(\alpha-1)[1+\xi^2/\Psi] - \varepsilon[1-\xi^2/\Psi]\right\}^2 \\ & -\mu\left\{\tfrac{\tau\gamma(\alpha-1)\xi - \Delta s\gamma(1+\alpha)\xi + \xi\varepsilon}{\Psi}\right\}^2. \end{aligned} \tag{13}$$

3.3. *Politically feasible free trade*

This section considers, analogous to Grossman and Helpman (1993), under which conditions the home country is willing to move towards free trade and to give up the instrument of tariffs to cushion external shocks.[10] It is assumed that the country enters a free trade arrangement or customs union (for the differences from a political economy perspective see Krueger, 1995) which is binding and cannot be left at discretion. In doing so, the government forgoes the possibility of reacting to external shocks which might hit the import-competing sector. Hence the expected size of the shock ~~and/or~~ the probability that a shock will occur in the foreseeable future enters the considerations of the government. Given the possibility that a shock, denoted by (1–P), occurs the government will enter the arrangement only if $\mathrm{EV}_F > \mathrm{EV}_N$, where E signifies expected values. That means that expected political support under free trade must be higher than expected utility under the case of non-retaliation. Given the short-time orientation of the government, which usually only covers the period until the next election, a government might be able to assess the probability of a shock for that rather short period. This does, of course, not exclude that the government might be surprised and be faced with political support other than expected.

The government will sign a free trade agreement if political support under free trade, given the probability of the shock, is larger than political support under non-retaliation and protection. This is the case if:

$$
\begin{aligned}
&\left\{\tau\gamma^3(\alpha-1)-\Delta s\gamma^3(1+\alpha)+\gamma^2\varepsilon-\alpha\gamma\mu t^*\right\}^2\\
&+\mu\left\{\tau\gamma^2(\alpha-1)-\Delta s\gamma^2(1+\alpha)+\gamma\varepsilon+\alpha\gamma^2 t^*\right\}^2\\
&\geq P(\gamma\varepsilon)^2(\gamma^2+\mu).
\end{aligned}
\tag{14}
$$

This consideration changes when the setting of trade policy is subject to retaliation. Again the government compares political support in the situation of free trade and discretionary tariff setting. Then the necessary condition for free trade to be viable is $\mathrm{EV}_F > \mathrm{EV}_R$.

This is given if political support under free trade is larger than political support under retaliation. For the given probability of a shock, this requires:

$$
\begin{aligned}
&\left\{\tau\gamma\xi^2(\alpha-1)-\Delta s\gamma\xi^2(1+\alpha)+\xi^2\varepsilon\right\}^2\\
&+\mu\left\{\tau\gamma\xi(\alpha-1)-\Delta s\gamma\xi(1+\alpha)+\xi\varepsilon\right\}^2\\
&\geq P(\xi\varepsilon)^2(\xi^2+\mu).
\end{aligned}
\tag{15}
$$

The comparison of the two cases, equations (14) and (15), makes clear that the degree of retaliation via $\xi \equiv \gamma(1-\alpha\theta)$ has a decisive influence of the political viability of free trade. The larger the degree of retaliation θ is, weighed with the size of the export sector α, the smaller the optimal tariff set by a political support maximizing government is. α being large, it is relatively easy for the government to give up the tariff instrument and the chances for free trade increase. In both cases social integration $s^* = s$, or preferably an even higher social standard abroad $\Delta s > 0$, can substitute tariffs and make free trade viable.

4. Conclusion

In this paper it is argued that two factors, which are mostly disregarded in the literature, play a pivotal role in determining whether a government enters a free trade arrangement or not. Geographical distance, reflected in transportation costs, creates a conflict between import competing and exporting industries, as does the setting of tariff rates. Exporters would prefer integration, because they suffer from tariff rates abroad. Moreover, they prefer integration with countries which are nearby because less transportation costs increase their competitiveness abroad. For the same reason, import-competing industries see transportation costs as a natural shelter against competitive pressure from abroad.

Both industries in a given country, however, have the same position concerning labor standards. They prefer to see higher standards abroad than in the home country because this increases their competitiveness. Starting out from a position of higher standards at home, it is obvious that both industries will combine their efforts to reach a political position which will force foreign governments to raise their labor standards as well, thereby raising the "rival's costs".

Changes of trade policy over time can therefore be explained by changes in the relative political influence of the sectors considered. The other important factors are, of course, a change in the degree of retaliation, leading to lower tariffs under higher retaliation, and a leveling of social standards. A redistribution across countries could also considerably change the optimal rate of tariff. The EU with its regional cohesion funds might be a good example of how those are used as a side-payment for diminishing the social divergence in the member countries.

Following this theoretical treatment, an empirical test should follow. The structure of an economy, i.e. the size of the export and importing sectors, its competitiveness, reflected in the current account, and its level of social standards should be directly related to its position in free trade negotiations. Countries with higher standards should only be willing to integrate when others raise their social standards as well. The negotiations about the social protocol in the EU indicate that this is in fact the case. If export interests are dominant, and if countries are internationally competitive, it should also follow that free trade areas are regionally concentrated (which is often but not always the case). To test this reasoning, the EU, NAFTA and other regional arrangements should be studied carefully.

Notes

1. This disregards the general equilibrium effects in a neoclassical model. In such a world exporters have an interest in trade policy because factor prices are affected (see e.g. Findlay and Wellisz, 1983). In the real world these effects are likely to play no decisive role.
2. A different view is presented by Bagwell and Staiger (1990) who argue that the chances for protectionist tariffs are higher in times of high volumes of trade because countries want to compensate for perceived terms-of-trade losses through increasing their tariffs.
3. Under cross-retaliation other domestic industries would lobby to avoid protection for an import-competing industry.
4. This interpretation fails to explain the Swiss and Norwegian decisions not to join the European Union (EU), leaving Switzerland and Norway the only countries in Western Europe outside the EU. One reason for this outcome might be that general referenda were held.
5. In a related concept when discussing monetary union in Europe, Alesina and Grilli (1993) have pointed out that two speed monetary union might prompt the countries inside monetary union to reject outside countries and their later accession to the union.
6. Frankel and Wei (1995) show that much of the high level and increase in intra-European trade is explained by size, GDP/capita, proximity and common borders of European countries.

7. Lange (1993) likewise points out that standard setting through the community can become an instrument in international competition policy.
8. Other fields of common regulation to lower competition are environmental regulation, competition policy and taxation (Vaubel, 1995).
9. This would of course be a redistribution in the higher developed country. While the trading industries would benefit from social cohesion, the costs of regional development funds would be borne by all factors of production through increased taxation.
10. In general a cooperative tariff could be chosen in negotiations which is lower than the tariff set before, but not necessarily zero. For simplicity only the cooperative tariff is set to zero here.

References

Alesina, A. and Grilli, V. (1993). On the feasibility of a one-speed or multi-speed European Monetary Union. *Economics and Politics* 5: 145–165.

Bagwell, K. and Staiger, R. (1990). A theory of managed trade. *American Economic Review* 80: 779–795.

Baldwin, R. (1993). A domino theory of regionalism, NBER working paper 4465, September.

Bergstrand, J.H. (1985). The gravitiy equation in international trade: some microeconomic foundations and empirical evidence. *Review of Economics and Statistics* 67: 474–481.

Corden, W.M. (1984). Normative trade theory. In: R. Jones and P. Kenen (Eds.), *Handbook of international economics* Vol 1, Chapter 2. Amsterdam: North-Holland.

Eaton, J. and Grossman, G.M. (1985). Tariffs as insurance: Optimal commercial policy when domestic markets are incomplete. *Canadian Journal of Economics* 18: 258–272.

Findlay, R. and Wellisz, S. (1983). Some aspects of the political economy of trade restrictions. *Kyklos* 36: 469–481.

Frankel, J. and Wei, S.-J. (1995). European integration and the regionalization of world trade and currencies: the economics and the politics, mimeo. University of California at Berkeley.

Grossman, G.M. and Helpman, E. (1993). The politics of free trade arrangements, NBER working paper 4596, December.

Hillman, A.L. (1982). Declining industries and political-support protectionist motives. *American Economic Review* 72: 1180–1187.

Hillman, A.L. and Moser, P. (1995). Trade liberalization as politically optimal exchange of market access, In: M. Canzoneri, W. Ethier and V. Grilli (Eds.), *The new transatlantic economy*, forthcoming.

Hillman, A.L. and Ursprung, H.W. (1988). Domestic politics, foreign interests and international trade policy. *American Economic Review* 78: 729–745.

Krueger, A.O. (1995). Free trade arrangements versus customs unions, mimeo. Stanford University.

Lange, P. (1993). Maastricht and the social protocol: Why did they do it?. *Politics & Society* 21: 5–36.

Magee, S.P., Brock, W.A. and Young, L. (1989). *Black hole tariffs and endogenous policy theory*. Cambridge: Cambridge University Press.

Mc Keown, T. (1986). The limitations of 'structural' theories of commercial policy. *International Organization* 40: 43–64.

Moser, P. (1990). *The political economy of the GATT*. Gruesch: Ruegger.

Rodrik, D. (1994). What does the political economy literature on trade policy (not) tell us that we ought to know?. NBER working paper 4870, September.

Salop, S. and Scheffman, T.D. (1983). Raising rival's costs. *American Economic Review* 73 (Papers and Proceedings): 267–271.

Schuknecht, L. (1992). The political economy of current European integration, In: H.-J. Vosgerau (Ed.), *European integration in the world economy*, 677–702. Berlin: Springer.

Sinn, G. and Sinn, H.-W. (1993). *Kaltstart: Volkswirtschaftliche Aspekte der deutschen Vereinigung*. München: Beck.

Vaubel, R. (1995). Social regulation and market integration: A critique and public choice analysis of the Social Chapter, mimeo. University of Mannheim.

42

WHARTON REPROGRAPHICS

The Causes of Regionalism

Richard E. Baldwin

LIPPINCOTT LIBRARY
DOCUMENT DELIVERY

1. INTRODUCTION

THE rapid spread of regionalism is surely the most important recent development in the global trade system. The big question — 'Is regionalism bad?' — has three parts.

(1) Do existing regional liberalisation efforts harm living standards in participating or nonparticipating countries?
(2) To date, has regionalism harmed the world's trading system and hindered multilateral trade liberalisation?
(3) Is regionalism likely to damage the global trade system in the future?

Economists who are familiar with the literature largely agree on the first part's answer. Almost all empirical studies of European and North American arrangements find positive impacts on members' living standards and inconsequential impacts on nonmembers' living standards. Empirical work on smaller arrangements is scarce, but there is little evidence that *bona fide* regional liberalisation has significantly lowered the living standard of any nation.[1]

There is less agreement on how to answer the second part, but history suggests a 'No'. Multilateral liberalisation since WWII has been wildly successful at cutting developed nations' industrial tariffs. Yet the nations that steered this multilateral liberalisation — the US, the UK, the EEC6, the Nordics and Canada — are the same ones that drove regional liberalisation since the 1958 Treaty of Rome, the 1960 Stockholm Convention, and the 1965 Canada-US Auto Pact.[2] Moreover, where GATT Rounds failed, the regional route also failed.

RICHARD BALDWIN is at the Graduate Institute of International Studies, Geneva. He thanks Andre Sapir, Bernard Hoekman, Alan Winters, Bob Baldwin, and a referee for comments.

[1] Yeats (1997) is almost an exception. He found that Mercosur's trade rose in sectors where its members do not (according to him) have a comparative advantage. If Mercosur's economy was perfectly competitive with constant returns, this would suggest negative welfare implications. He explicitly avoids discussing welfare, however, since he cannot measure potentially positive scale, procompetitive, variety, or growth effects. See Baldwin and Venables (1995) for the economics of regional arrangements and a survey of relevant empirical studies. Finally, some arrangements in Africa raised rather than lowered intra-bloc barriers. Results for such groups tell us nothing about the welfare impact of preferential *liberalisation*.

[2] GATT stalwarts Australia and New Zealand avoided regionalism until recently.

© Blackwell Publishers Ltd 1997, 108 Cowley Road, Oxford OX4 1JF, UK
and 350 Main Street, Malden, MA 02148, USA.

WORLD ECONOMY 20 (7): 865-888.

Historically, developed nations resisted agricultural trade liberalisation both multilaterally and regionally, and developing nations resisted tariff-cuts on manufactures both multilaterally and regionally.[3] Finally, at least one regional integration effort promoted multilateral liberalisation via 'competitive liberalisation' (to use Fred Bergsten's recent term). Most scholars believe that formation of the EEC induced the US to push for multilateral tariff-cutting via the Kennedy Round. Kenen (1989) provides the textbook treatment, which entails two steps. First, the phase-in of EEC preferences harmed US exporters. Second, redressing the discrimination required multilateral liberalisation (the direct way to offset it, joining the EEC, was unthinkable). In this roundabout way, European regionalism created a burst of free-trade multilateralism among US exporters, and the US government followed suite.

The third part of the question is the hardest and here economists seriously disagree. The Larry Summers school-of-thought (Summers, 1991) looks at regional, i.e. discriminatory, liberalisation and sees liberalisation. An outcome it welcomes since it views regionalism as having a largely benign effect on the multilateral system. The Jagdish Bhagwati school (Bhagwati and Krueger, 1995) looks at discriminatory liberalisation and sees discrimination. An outcome it views as a serious threat to the WTO-centred world trading system. The risks it sees include: inter-bloc trade war, greater dominance of small nations by hegemonic powers, and dampened enthusiasm for further multilateral liberalisation.

I do not know how to definitely judge which school more accurately predicts the future. But I am sure that the first step is to look to the past. That is, to understand the future consequences of regionalism, one must determine the causes of past regionalism.

There is a mainstream explanation of the causes of recent regionalism.[4] It rests on two pillars: (1) frustration-with-GATT/regionalism-is-easier, and (2) the United States's conversion from devoted multilateralist to ardent regionalist. If this explanation is right, regionalism is indeed a dangerous drift and strong medicine is called for. Regional initiatives substituted for multilateral liberalisation and the world's biggest player is disaffected with the system.

I will argue that the traditional explanation is inconsistent with the facts. I will also present an alternate reading of recent history based on a domino theory of regionalism. At this point, however, a quotation from the great sleuth is in order — Holmes to Watson in The Hounds of Baskerville:

> It is a capital mistake to theorise before one has all the facts. Inevitably one twists the facts to fit the theory, rather than the theory to fit the facts.

---

[3] Food trade is excluded in most preferential arrangements. The EU has open trade in food, but the CAP hardly constitutes free trade. South-South arrangements are often not implemented (De Torre and Kelly, 1992; and Schott, 1989), and developing nations rarely cut tariffs in GATT Rounds.

[4] This characterisation draws on Bhagwati (1993), Krugman (1993) and Panagariya (1966), *inter alios*. Particular authors, of course, differ on many details.

© Blackwell Publishers Ltd 1997

Taking this advice, Section 2 examines the facts. Section 3 describes and critiques the standard explanation. Section 4 argues that a domino theory of regionalism better fits the facts. Lastly, Section 5 contains a summary and speculation on the consequences of regionalism.

2. RECENT REGIONALISM: THE FACTS

a. The Importance of Europe and North America

Spreading regionalism is on the world policy agenda because many respected thinkers fear that it could kill the proverbial gold-laying goose — the world trading system. The fear-inducing regionalism, however, does not consist of the Andean trade group, the East African Economic Community, or other similar arrangements. Members of such clubs are economically small, account for very little of the world's trade, and have more frequently been free-riders than supporters during GATT Rounds. North Atlantic regionalism is an entirely different matter.

European and North American regional arrangements directly govern 40 per cent of world trade, and their members account for half of world income. These agreements affect trade with non-member nations, so it is important that two-thirds of world trade is with North Atlantic nations that have discriminatory deals with their main commercial partners. Furthermore, these North Atlantic nations have been the traditional stalwarts of the WTO/GATT. Plainly, the goose is in trouble if regionalism greatly weakens Europeans' and Americans' support of multilateralism. It is therefore paramount to understand their recent interest in regionalism.

Regionalism in Asia (by itself or with the Pacific nations) would certainly be an important event, but has yet to happen.[5] Nor is it likely to happen soon according to Lawrence (1966, Chapter 7). Martin Wolf makes the point succinctly:

> By now, no fewer than 85 regional trade arrangements are in existence, 28 of which have been set up since 1992 (see GATT, 1993). Yet the world has only two significant trade blocs, defined as zones of preferential trade. One is western Europe and the second is north America. ... Meanwhile, Asia is not a bloc and shows little sign of becoming one (Wolf, 1994, p. 13).

b. Two Types of Recent Regionalism: Deep and Shallow Arrangements

International integration agreements are contracts that restrict governments' ability to set policy unilaterally. Some regional integration schemes are shallow;

---

[5] The Asia-Pacific Economic Cooperation (APEC) forum covers a quarter of world trade, but is not a preferential arrangement and virtually no APEC-specific liberalisation has occurred to date (Bergsten, 1997). The ASEAN free trade area, which covers only 1.3 per cent of world trade, was agreed in 1992, but has made little progress.

© Blackwell Publishers Ltd 1997

others are deep. The distinction matters in Section 3, so I cover the relevant facts here.

(i) Deep integration schemes.

The EU's Single Market (EC92) programme and the Maastricht Treaty are archetypical deep integration agreements. In principle, they mean that member states cannot restrict imports of EU goods or services at the border (all border measures including anti-dumping and countervailing duties are excluded), or by means of domestic health, safety and environmental standards (standards are mutually recognised). Nor can they restrict intra-EU capital and labour movements, or inhibit establishment of firms from EU nations. Member states cannot freely choose their state-aids policy, competition policy, or indirect taxation (VAT) rates. All these restrictions are supervised by the supranational European Commission and can be enforced by the supranational European Court of Justice (this can overrule member state courts on Single Market issues). Lastly, nations joining the prospective monetary union will no longer control their monetary policy. Of course, practice does not always match principle, but obviously many of these policies would be considered matters of purely national concern in, say, North America and Japan. This description also explains why so few deep-type regional arrangements exist.

Except the Australia-New Zealand Closer Economic Relations Trade Agreement (ANZCERTA), all deep integration schemes are in Europe. For example, the European Economic Area (EEA) agreement extends single-market access to Norway, Iceland and Liechtenstein, and the so-called Europe Agreements commit ten Central and Eastern European countries to approximate their economic policy laws and practices to the EU standard.

(ii) Shallow regional integrations.

Most recent regional schemes are free trade agreements (FTAs). The list includes the US-Canada FTA (CUSFTA), and US-Israel FTA, the North American FTA (NAFTA), Mercosur, and most other regional schemes in Latin America, Africa and Asia. These involve shallow integration, often little more than tariff and quota liberalisation. Two agreements are worth looking at more closely. The US-Canada FTA eliminates all bilateral tariffs by 1998, and relaxes some restrictions on government procurement and cross-border direct investments. Service trade is not substantially liberalised (most existing restrictions are 'grandfathered'). Many complex issues — most notably subsidies, energy sector investment, transport services, and intellectual property rights — proved insurmountable in the bilateral talks and were dropped (Hufbauer, Schott and Clark, 1994, p. 99).

NAFTA will eliminate most tariffs and many non-tariff barriers on US-Mexico and Canada-Mexico trade in 10 to 15 years (CUSFTA still governs US-Canada trade). It liberalises many investment restrictions among the three, and taken together, the investment component of NAFTA goes beyond those agreed in the

© Blackwell Publishers Ltd 1997

Uruguay Round. Trade in financial and telecommunications services is also partially liberalised. Finally, NAFTA is not a customs union, so rules-of-origin are necessary. Krueger (1993) argues that these are *de facto* protection, a point that also applies to Europe Agreements and the EEA.

c. The Timing and Motivation of Regional Arrangements

The timing and motives are critical — yet open to interpretation — so I cite well-known trade policy experts on the key facts.

(i) North American regionalism.

The US and Canada have a long history of bilateral trade talks, with major deals struck, or almost struck in 1854, 1874, and 1911. Yet ever since the GATT was signed, the US has almost continuously sought to liberalise bilateral trade. In March 1948, the two nations concluded a secret draft protocol eliminating most tariffs and quotas bilaterally, but this was ultimately rejected by the Canadians (Smith, 1988, p. 39; and Wonnacott, 1987, p. 15). In 1958 US government procurement was preferentially liberalised in Canada's favour by the Defense Production Sharing Agreement (this was linked to Canadian purchases of US arms). In 1965 the Canada-US Auto Pact was signed, granting duty-free treatment to a wide swath of bilateral commerce, namely trade in automobiles and parts.

After failure of the 1948 effort, US interest in an across-the-board free trade agreement with Canada only re-emerged in 1974, when Section 612 of the Trade Act of 1974 authorised the US President to negotiate an FTA with Canada. Interestingly, this came soon after an ambitious agenda had been agreed for the Tokyo Round. Also just as the Tokyo Round was winding up, Congress further probed the possibility of a CUSFTA. The Trade Agreements Act of 1979, section 1104, required the President to study 'the desirability of entering into trade agreements with countries in the northern portion of the western hemisphere' (Wonnacott, 1987, p. 19). The resulting report to Congress, USTR (1981), found that the Auto Pact had worked well and that 'further opportunities to rationalise industries through freer trade should be explored.' According to Wonnacott, the governments agreed upon four sectors for negotiation, but talks never began. At about the same time, Ronald Reagan espoused North American free trade in his 1980 Presidential campaign.

This long-running US interest in regionalism was rejected by Canada until the mid-1980s (Whalley, 1993, p. 369; Schott, 1987, p. 16; and Hufbauer, Schott and Clark, 1994, p. 98). An FTA had long been the *bête noir* of Canadian politics, pitting commercial export interests against Canadian fears of economic/cultural domination by its giant neighbour.[6] Ottawa's policy, in fact, systematically

---

[6] For example, Wonnacott (1987) says that Canadian premiers Diefenbaker and Trudeau made popularist pledges to reduce Canadian trade dependence on the US in the 1950s and 1970s.

© Blackwell Publishers Ltd 1997

deterred US involvement in Canadian industry, natural resources and media. A major factor in Canada's conversion to the FTA idea was the 1980s rise in US protectionism (Smith, 1988, p. 41; Wonnacott, 1987, p. 17; and Schott, 1988, p. 29). Indeed, countervailing duties and other unfair trade remedies were a central Canadian concern in CUSFTA talks. The Canada-US FTA started formally in September 1985 with Prime Minister Mulroney's proposal (President Reagan accepted in December 1985). Despite their economic and institutional similarity, the success of the 1965 Auto Pact, and the low initial level of US and Canadian tariffs, the talks proved difficult (Bello and Winham, 1992). CUSFTA entered into force only in January 1989.

Mexico was the driving force behind the US-Mexico FTA (Hufbauer, Schott and Clark, 1994, p. 100; and Whalley, 1993). Until the 1980s, Mexico pursued protectionist and anti-foreign investment policies. Its early 1980s' debt crisis, however, precipitated a Mexican *volte-face*, which included the adoption of a liberal trade policy. Mexico unilaterally cut its tariffs and non-tariff barriers (on a non-discriminatory basis), joined the GATT, and signed several bilateral trade accords with the US and Canada.[7] The process culminated with the Mexican President's June 1990 proposal for an FTA with the US.[8] According to Whalley (1993), Mexican objectives were to improve and lock-in access to the US market, and to underpin its domestic economic reforms with the aim of attracting foreign direct investment.[9] US President Bush rapidly agreed to bilateral talks. The US motives had little to do with trade liberalisation. The chief US aim was to foster stability in Mexico by boosting growth and locking in unilateral pro-market reforms. While the motives for the US-Mexico FTA stemmed from Mexico's idiosyncratic *volte-face*, the motives for NAFTA involved predictable forces.

The initiative to turn the US-Mexico talks into NAFTA came solely from Canada, which feared the possible trade diversionary effects of a US-Mexico FTA.[10] Thus just 18 months after the politically-bruising US-Canada deal, Canada requested that the US-Mexico talks to be trilateralised. After promising not to hinder a US-Mexico accord, Canada's request was accepted in February 1991. Talks began in June 1991 and finished in December 1992. NAFTA took effect in January 1994.

(ii) Knock-on effects in Latin America.

In Latin America, NAFTA — especially its US-Mexico trade provisions —

---

[7] Understanding on Subsidies and Countervailing Duties 1985, Framework of Principles and Procedures for Consultations Regarding Trade and Investment, 1987, and Understanding Regarding Trade and Investment Facilitation Talks 1989. Ten agreements were signed with Canada in March 1990.

[8] Schott (1989, Chapter 1) details traditional Mexican resistance to a US-Mexico FTA.

[9] See Fernandez (1997) on the 'lock-in' claim, and Blomstrom and Kokko (1997) on foreign investment effects.

[10] See Schott and Hufbauer (1992, p. 63) for discussion and quantitative estimates.

© Blackwell Publishers Ltd 1997

created what Lawrence (1996, p. 77) calls 'powerful pressures for inclusion'. Chile, Brazil, Argentina, Uruguay and Paraguay all formally or informally approached the US with requests for FTAs. The Bush administration, fearing five separate battles with protectionists in Congress, discouraged these efforts and offered the Enterprise for the Americas Initiative (EAI) instead (this was launched in June 1990). According to CEA (1991, p. 255), the EAI created a process leading to free trade agreements, the first step of which required nations to sign so-called Framework Agreements. These committed countries to reducing their investment restrictions, *inter alia*, in exchange for promises of closer US ties in the future. Although the Framework Agreements might be viewed as unilateral 'concessions' by merchantilists, 26 countries signed them in 1991.

Serra et al. (1977, p. 8) also claim that NAFTA 'stimulated a series of overlapping bilateral and plurilateral agreements in the Hemisphere.' For instance, Mexico signed FTAs with Chile, Caricom (a Caribbean trade group), Colombia, Venezuela, Costa Rica and Bolivia. Chile signed a similar list of FTAs, the Andean Pact (Venezuela, Columbia, Ecuador, Peru and Bolivia) was renewed, and Mercosur (a customs union between Argentina, Brazil, Paraguay and Uruguay) was established in March 1991.

The most important of these is Mercosur. The connection between NAFTA and Mercosur is difficult to pin down, but in 1990 Bush administration trade officials suggested that the Southern Cone countries band together in their own regional arrangement before pursuing an FTA with the US. It seems unlikely that such a suggestion was the prime cause of Mercosur (Latin American affinity for trade blocs goes way back). However, Mercosur was launched after US unresponsiveness closed the bilateral route to the US market.

Chile resisted Mercosur membership until October 1996, and even then only became an associate member. There may be many reasons for Chile's timing, but two ingredients stand out. First, while Chile headed the queue for a US bilateral, it was clear by 1996 that the US Congress was not in the mood for such talks. Second, Mercosur started to create its own 'pressures for inclusion'. Unlike most South-South FTAs, Mercosur's preferential liberalisation was implemented and did affect trade patterns. From 1990 to 1995, intra-Mercosur trade expanded at 27 per cent per year while external trade expanded at only 7.5 per cent (Yeats, 1997). Clearly the arrangement favoured within-block firms relative to extra-bloc firms.

The manifest 'success' of Mercosur continues to exert a gravitational pull. After Chile signed up in October 1996, Bolivia became an associate member in December 1996. It has even attracted the attention of the EU. According to the *Financial Times* (4 February, 1997, p. 12), the EU is engaged in slow-going talks to establish an EU-Mercosur FTA.

Finally, Jamaica and Trinidad and Tobago applied for NAFTA membership in 1995. According to Commonwealth Secretariat (1995), NAFTA provisions for increasing Mexico's sugar and clothing exports to the US threaten these sectors in

Jamaica, and Trinidad and Tobago. A threat that helps to explain the timing of the pair's application.

(iii) European regionalism.

European integration has a long history, but the most recent wave was sparked by the Single Market programme, EC92. In 1985 Jacques Delors took office (as President of the European Commission) with the intention of rekindling European integration via completion of the internal market. The first step, the 1985 White Paper by Lord Cockfield, recommended changes that were eventually implemented by the 1986 Single European Act. In the same year, Spain and Portugal acceded to the EU after eight years of negotiations.

The Single Market programme harmed EFTA-based firms by lowering the costs of their EU-based rivals, as Krugman (1984) and CEPR (1992) point out. EFTA governments decided to react in 1984, even before the final draft of Cockfield's White Paper. However, the intrinsic complexity of single-market issues and the EU's preoccupation with monetary union led to long delays. Talks were only concluded (after three renegotiations) in 1993. Granting single-market access requires supranational surveillance and enforcement. EU members accept this since they oversee the process politically. Non-member governments, however, cannot vote in or attend the key EU meetings, so the EEA agreement entails some regulation without representation. This feature made the EEA relatively unpalatable. In fact almost none of the EFTA governments could live with it. All EFTAns, except Iceland and Liechtenstein, applied for EU membership *before* the EEA talks concluded. The Icelandic government, which gave much thought to joining in the early 1990s, was initially deterred by the EU's common fishery policy and its low dependence on the EEA's manufactured goods market. Iceland changed its mind in 1994, but was told that this was too late for the 1995 enlargement. Iceland therefore withheld its application. The applications led to accession talks, which concluded in 1994. Enlargement was to occur in 1995 subject to national referendums. Austrian, Finnish and Swedish voters responded positively. The Norwegians said no (Swiss voters had earlier rejected even the EEA).

Note that West European deep integration initiatives (EC92, Maastricht and EEA) did not expand regional integration, they merely deepened it. Before 1986, West European regional integration was marked by two concentric circles. The outer circle (EU plus EFTA members) was a 'virtual' free trade area for industrial goods due to the confluence of the Treaty of Rome, the Stockholm Convention, and the bilateral EU-EFTA FTAs. The inner circle was the EU's common market. EC92 and the EEA deepened integration in both circles, but the 1986 and 1995 enlargements brought only one new country (Spain) into the two-circle system.

An important geographical expansion of European regionalism followed the collapse of communism. In the 1990s, the EU and EFTA separately signed

© Blackwell Publishers Ltd 1997

bilateral free trade agreements with ten Central and Eastern European countries (CEECs).[11] All ten applied for EU membership and have been assured of eventual admittance. Geopolitical considerations were the main motives for this expansion of regionalism. Even more recently, the EU has signed 'Euro-Med' agreements with Tunisia and Morocco. The Euro-med pacts are industrial free trade agreements aimed at offsetting the preference-dilution that these two former French colonies suffered from the Europe Agreements (see Gala and Hoekman, 1997).

(iv) Asian regionalism.

ASEAN's first attempt at regional integration — its 1979 PTA — failed, according to DeRosa (1995), due to heavy reliance on administrative protection, and opposition by industrial and agricultural vested interests. ASEAN's recent attempt — its 1992 ASEAN FTA (AFTA) — may suffer the same fate. AFTA is politically and economically a heterogeneous group, and intra-AFTA trade accounts for only 15 per cent of AFTA exports, about half of which is entrepot trade with Singapore that is basically duty-free already, according to DeRosa (1995). Additionally, AFTA members vary greatly in the composition of their exports (industrial versus agriculture goods, for instance) and in their current levels of protection. Consequently, going to zero tariffs on substantially all goods will entail very asymmetric — and therefore politically difficult — 'concessions' in terms of both tariff-cutting and 'winning and losing' sectors.

The APEC initiative, launched in November 1994 in Jakarta, has consciously avoided commitments to preferential liberalisation for several reasons. For instance, the article 24 route would require them to adopt a common plan for achieving zero-tariff intra-bloc trade on substantially all goods within a fairly short period.[12] Many APEC nations — Malaysia for example — are not prepared to alter state-led development plans for the sake of free trade. Moreover, they fear domination by the US and/or Japan.[13] APEC consequently relies entirely on unilateral, MFN liberalisation (called Individual Action Plans), and has adopted a very long-term goal for free trade: 2020 for most, 2010 for developed members. This raises doubts about APEC's credibility since, as Guy de Jonquieres unkindly remarks in his 18 November, 1994, *Financial Times* column: 'By then, most of the leaders who met in Jakarta will be out of office, forgotten or dead.'

Lawrence (1996) provides four reasons why APEC is unlikely to become a trade bloc soon. First, Asia has so far prospered without regionalism and as the

---

[11] Poland, Hungary, Czechia, Slovakia, Bulgaria, Romania, Estonia, Latvia, Lithuania, and Slovenia.

[12] GATT Article 24 sanctions and regulates FTAs and customs unions.

[13] See Guy de Jonquieres' opinion piece (*Financial Times*, 18 November, 1994) entitled 'Different aims, common cause' on the lattice of motives driving APEC.

© Blackwell Publishers Ltd 1997

old saying goes: 'If it ain't broke, don't fix it.' Second, given its economic size, Japan would inevitably play a major role. Memories of Japan's 1930s efforts at achieving regional hegemony make this an unappealing prospect to today's Japanese and other Asians.[14] Third, the threat of exclusion, which loomed large in the motives behind European and North American regionalism, is largely absent in Asia. Developing Asian economies do not depend upon Japan as much as, say, Mexico and Canada depend on the US. Moreover, Japan has become more open, rather than less, in recent years. Finally, as mentioned above, many Asian nations just do not believe in free trade.

3. THE STANDARD EXPLANATION

The standard explanation of regionalism's recent resurgence focuses on two assertions. Paul Krugman phases the first one as a series of questions:

> So we need to ask why such blocs are in fact emerging. This comes down to asking why nations may feel that they are able to negotiate more at a regional than at a global level. Or to put it more pessimistically, what are the problems of the GATT that lead countries to turn to their neighbourhood instead? (Krugman, 1993, p. 73).

Krugman lists four reasons why regionalism is easier. (i) The large number of participants in multilateral trade negotiations reduces the costs of non-cooperation (a key consideration in cooperative game theory) and fosters greater rigidity in negotiations. (ii) Modern trade barriers are more varied, more complex and less transparent than the tariffs and quotas considered in the early GATT Rounds. This makes multilateral bargaining harder and renders monitoring of any eventual agreements more problematic. (iii) The reduced economic dominance of the US makes the world trade system harder to run, as suggested by the theory of hegemonic stability. (iv) Socio-politico-economic differences between Atlantic economies and Asia (Japan in particular) make it much easier for the latter to get away with non-compliance, so Atlantic economies are reluctant to bargain on an unconditional MFN basis. A slight variant of the first assertion is that regionalism spread due to frustration with the GATT in general and the Uruguay Round talks in particular.

Bhagwati (1993) states the second assertion crisply:

> The main driving force for regionalism today is the conversion of the United States, hitherto an abstaining party, to Article XXIV.

Some analysts — e.g. Bhagwati (1993), Panagariya (1996) and Bergsten (1996) — link the two assertions. For example, Bergsten (1996, p. 6) claims:

"The United States reversed its traditional aversion to regionalism by

---

[14] The importance of overseas Chinese in the conduct of intra-East Asian trade also matters given the resentment that this minority faces in many Asian nations.

© Blackwell Publishers Ltd 1997

embracing free trade agreements with Israel and Canada after the European Community blocked the launch of new negotiations in the GATT'' [in 1982].

a. Assertion No. 1 is Factually Incorrect

(i) Regional arrangements are not easier politically.

Much of the first assertion's plausibility stems from events of the early 1990s. While the Uruguay Round lurched from crisis to crisis, Europe and the Americas set up regional schemes at a mighty pace. From the post-Uruguay Round perspective, however, the regionalism-is-easier assertion is hard to justify (Cable, 1994). Canada's bilateral FTA with the US was a highly divisive issue in their subsequent parliamentary election, while the Uruguay Round enjoyed broad support. In the US, NAFTA faced much stiffer opposition than the Uruguay Round (witness the 1992 Presidential election debate). In Europe the picture is mixed. The Uruguay Round's impact on the Common Agricultural Policy brought the EU to the brink, while EC92 skated through the approval process. The EU's other main regional initiative, the Maastricht Treaty, created enormous political problems. In Asia, the Uruguay Round proved popular, but regional integration schemes have singularly failed to catch on.

(ii) Most regional schemes deal with tariffs and quotas.

Krugman's second reason certainly sounds right. Contemporary trade issues are more complex than those dealt with in the early GATT rounds. Krugman's theoretical point is, however, irrelevant to the regionalism that most concerns the Jagdish-Bhagwati-school, namely FTAs outside Western Europe. My point is simple. Shallow regional arrangements — essentially all non-European schemes — typically fail to address complex trade barriers, sticking mainly to tariffs and quotas. In the areas where these arrangements go beyond tariffs and quotas, the Uruguay Round also proved successful. WTO (1995, p. 62) points out that the Uruguay Round results went further down the liberalisation road than most regional integration agreements in the areas of services, investment, intellectual property rights, agriculture, clothing and textiles, government procurement, and the establishment of a credible dispute settlement process.

Krugman's notion is clearly relevant to deep European integration, yet two caveats apply. First, the most important revitalisation of European regionalism — EC92 and Maastricht — did not extend regionalism beyond its traditional West European borders. And even the Bhagwati-school views West European regionalism as unthreatening to the global trading system (Bhagwati and Krueger, 1995). Second, the regionalism-is-easier assertion does not explain the novel — and potentially threatening — aspect of European integration, namely its geographic spread. The Europe Agreements were driven by geopolitical factors, and the Euro-Med pacts involved only shallow integration.

To sum up, these two ideas — that regionalism was politically easier, and that GATT was unable to deal with complex trade barriers — are just plain wrong.

(iii) Frustration-with-the-GATT mattered little in North America and Europe. The notion that the US was transformed from a staunch multilateralist into a regionalist by GATT-frustrations appeared as early as Schott (1989, p. 4):

> As a result of the 1982 meeting, US Trade Representative William E. Brock also left Geneva with a bad taste for GATT negotiations, and soon after began to pursue bilateral talks with Israel and Canada.

Here again the facts just do not fit. Consider the US-Canada FTA.

A careful reading of history shows that the US was almost constantly interested in preferential trade with its northern neighbour.[15] This should be expected. While a Trade Representative's personal opinion occasionally matters, US trade policy is generally determined by the intricate interplay of US geopolitical goals and the demands of powerful, well-connected, well-financed special interest groups. US exporters are among the most powerful of these interests, and Canada is the number-one market for their goods. It is natural, therefore, that the US has always been interested in better access to the Canadian market. Turning from theory to evidence, here is the list of bilateral deals: an FTA protocol that Canada eventually rejected (1948), the Defense Production Sharing Agreement (1958), and the Auto Pact (1965). Moreover, in 1974 the US Congress (which does not delegate its constitutional powers lightly) granted negotiating authority for an FTA with Canada. In 1979 it mandated a study on North American FTAs. The resulting report to Congress, USTR (1981), recommended that further bilateral liberalisation with Canada and Mexico be explored. Notice that all of this was set in motion before the *Tokyo Round* was finished. It would take a very complicated story to explain how sluggish progress towards a new round in 1982, or later crises in the Uruguay Round prompted the US Congress (in 1974) and Brock's own office (in 1981) to favour a preferential trade with Canada. It would take an even taller tale to explain how European frustration with the GATT triggered recent European regionalism.

Quite simply, assertion No. 1 is factually incorrect.

b. Assertion No. 2 is Factually Incorrect

The second assertion — that a principal force behind recent regionalism is the conversion of the US from devoted multilateralist to ardent regionalist — is also based on a misguided reading of recent history. In the words of Hufbauer, Schott and Clark (1994, p. 98):

> While the political will for free trade negotiations was apparent in the United States as early as 1974, Canada was slow to welcome negotiations.

---

[15] See Wonnacott (1987), Smith (1988), or Hufbauer, Schott and Clark (1994).

© Blackwell Publishers Ltd 1997

Whalley (1993, p. 369) is more terse:

> It was Canada who initiated negotiations on CUSFTA, Mexico on NAFTA.

When it comes to other regional arrangements in the Western Hemisphere (the Enterprise for the Americas Initiative, etc.), the US-attitude-as-driving-force assertion is an unsatisfactory combination of tautology and ethnocentricity. Plainly, the US's willingness to engage in regional liberalisation is a necessary condition for agreements that involve the US. However, the remarkable point in the late 1980s and early 1990s was not the US's willingness to liberalise. What was surprising was the willingness of Latin Americans to lower their import and investment restrictions. Until ten years ago, Latin America resisted all forms of trade liberalisation — unilateral, regional and multilateral. It is surely rooster-makes-the-sun-rise reasoning to suggest that US attitudes towards regionalism account for Latin America's recent embrace of free markets and free trade.

Finally, it is even more implausible to assert that Mercosur can be explained by US policy makers' supposedly new-found love of regionalism. In fact, one reason that Mercosur was set up was the US's *unwillingness* to negotiate bilateral FTAs with the Southern Cone countries. Regardless, the spread of Mercosur — it has been enlarged to include Chile and Bolivia, and it has started trade talks with EU — can in no way be attributed to the US's conversion.

When it comes to Europe, the assertion becomes ridiculous. Europeans do worry about US attitudes when it comes to foreign and defence policy, and jealousy of US prowess is an important background force in European integration. However, anyone who follows European politics knows that the US attitudes towards regionalism had nothing to do with economic integration in Europe since 1986.

These lines of argument should be enough to reject assertion No. 2.

Having rebutted its two main assertions, I conclude the section simply. The standard explanation is appealing at first glance, but it is not consistent with the facts.

4. A DOMINO THEORY OF REGIONALISM

There are other explanations for spreading regionalism — Anderson and Blackhurst (1993), Whalley (1996), Lawrence (1996) and Bergsten (1996), for instance — but this is not the place to address them. Rather, I shall argue for a domino theory of regionalism (Baldwin, 1993). That is, idiosyncratic incidents of regionalism triggered a multiplier effect that knocked down bilateral import barriers like a row of dominos.

The basic idea is simple. Forming a preferential trade area, or deepening an existing one, produces trade and investment diversion. This diversion generates

© Blackwell Publishers Ltd 1997

new political economy forces in nonparticipating nations — what Lawrence (1996) calls 'pressures for inclusion'. The pressure increases with the size of the trade bloc, yet bloc size depends upon how many nations join. Clearly, then, a single incidence of regionalism may trigger several rounds of membership requests from nations that were previously happy as non-members. If the trade bloc is open to expansion, regionalism may spread like wildfire. If the enlargement 'burn-path' is barred, the new political economy flames may find vent in preferential arrangements among excluded nations. Notice that such regionalism could occur despite any progress being made in ongoing multilateral talks, unless these also promised to fully offset the discrimination.

The idea underlying the domino explanation is hardly new. A presentation can be discerned in Jacob Viner's account of how dozens of German principalities and city-states were cajoled and coerced into joining Prussia's Zollverien between 1819 and 1867 (Viner, 1950, Chapter V.3). I imagine, however, that the notion is much older; Bismarck himself probably understood the political-economy dynamics of trade diversion. More recently, Whalley (1993) also views Western Hemispheric regionalism as largely defensive (rather than US-led). He, however, focuses on fears of US protectionism instead of trade diversion, and he does not posit a circular causality between bloc size and the strength of inclusionary pressures.[16]

Be that as it may, it is important to minutely examine the internal logic of even the oldest ideas and this requires maths. To my knowledge, Baldwin (1993), which deals with EU enlargement, was the first formal framework to explore the dynamic political-economy implications of trade diversion. Several subsequent studies — most notably Yi (1996), Ethier (1996) and Siedmann and Winter (1997) — have elaborated the basic framework.[17] Fortunately, an idea is simple to describe in words (as I shall do here) once its internal logic has been established.

a. Details

Clarifying the domino logic requires a bit of background. National trade policies are endogenous. In a price-like manner, proposed trade policies 'clear' the political market by equalising the demand and supply of protection. This approach — see Hillman (1989) for the basic theory, and Dixit, Grossman and Helpman (1997) for recent refinements — also applies to national attitudes

---

[16] Whalley (1993, p. 369) writes that what Mexico, Canada and other small nations sought was:

> safe-haven agreements with larger trading powers, seen by them as helping them avoid being sideswiped by protectionists barrier increases by their larger trading partners even if directed at other countries (the United States against Japan, for instance).

[17] Winters (1996) surveys all 'regionalism vs multilateralism' models, putting the domino theory in perspective. Baldwin (1993) was published as Baldwin (1995).

© Blackwell Publishers Ltd 1997

towards regional groups. In particular, a nation's stance on regional integration balances anti-membership and pro-membership forces. Among the important pro-integration forces are national firms that export to the regional bloc or potential FTA partner.

Trade liberalisation, however, poses a puzzle for this approach. Why should a nation remove barriers that it previously found politically optimal to impose? As it turns out, new regional initiatives provide a natural resolution to the puzzle.

Consider the impact of deepening an existing bloc's integration or the creation of a new FTA. Since the new discrimination lowers non-member firms' profit, closer integration stimulates the exporters to engage in greater pro-integration political activity. If the government was close to indifferent to membership initially, the extra activity may tilt the balance, causing the country to join. If it does, the block enlarges and the expense of non-membership increases. In the case of an FTA, enlargement harms non-member firms by lowering the costs of a new set of their rivals. In the case of a customs union or single market, extension of the common external tariff may cause further harm. All this encourages a second-round of pro-integration political activity, which may further enlarge the bloc.

If the trading bloc is truly open, the process stops only if some outsiders have very high, intrinsic resistance to membership. Here the cases of Switzerland and Norway come to mind. If the block is not entirely open, the process may stop when incumbents refuse to admit an outsider whose membership would be very costly (economically or politically). Here the cases of Turkey and Slovakia come to mind. If the bloc is entirely closed, membership requests will pile up, and the outsiders may turn elsewhere. Here the case of NAFTA comes to mind.

(i) Asymmetric lobbying: losers lobby harder.

The force-for-inclusion is strengthened by a peculiar tendency of special interest groups. They fight harder to avoid losses than to secure gains. This is important when thinking about regional versus multilateral liberalisation. The discriminatory aspects of trade blocs actually harm non-member exporters. The same need not be true of a failed multilateral liberalisation, so exporters may lobby harder to redress discrimination than they would to ensure that a multilateral round succeeds.

While there may be many explanations for this asymmetric lobbying phenomenon, Baldwin (1993) proposes one based on sunk costs.[18] Entry into most industries and markets involves large unrecoverable investments in product development, training, brand-name advertisement and/or production capacity. Established firms can earn profits without attracting new firms only to the extent that the profit constitutes a fair return on entry investments (in short, sunk costs create quasi-rents). Consider the incentive to lobby in such industries.

---

[18] Available from Baldwin@hei.unige.ch. Grossman and Helpman (1996), and Baldwin and Baldwin (1996) extend the idea. For earlier work see Hillman (1982) and Baldwin (1982).

If a country's exporters obtain additional access to foreign markets (say, via a multilateral liberalisation), their sales and profits typically rise. The extra profit, however, attracts new competition, so the size of the gains must be limited, and in the extreme of free-entry and symmetric firms, they disappear. Correspondingly, the incentive to lobby for new export opportunities will be limited, and in the extreme will be non-existent. The firms, however, will react quite differently to an unanticipated, profit-reducing policy change (e.g. EC92 or the US-Mexico FTA). To be concrete, suppose firms' marginal costs are zero and the change wipes out half their quasi-rents. The firms will not shut down and, more to the point, they should be willing to spend much of the lost rents on pro-membership lobbying. If they cannot redress the discrimination at its source (i.e. membership is blocked), they may still find it profitable to lobby for multilateral liberalisation (think of US firms and the Kennedy Round), or for a new preferential arrangement (think of Mercosur). Notice that this latter statement is true, even if the firms had no interest in additional liberalisation before the unanticipated change.

b. American Dominos

The idiosyncratic event in the Western Hemisphere was the US-Mexico FTA.[19] Announcement of these talks faced other countries in the region, which are heavily dependent on the US market, with a *fait accompli*. Mexico-based producers would gain preferential access to the US market, thereby harming third-country exporters and diverting foreign investment to Mexico. Canada-Mexico tariffs were the first dominos to fall. Canada decided that it needed a place at the table to reduce possible trade and investment diversion (especially in auto parts). Thus Canada found itself lowering barriers that it had previously found optimal (politically) to maintain. More dominos would have fallen, had NAFTA been open to enlargement. Chile, Brazil, Argentina, Uruguay and Paraguay, which had previously found it politically optimal to maintain tariffs against US exports, offered to reciprocally lower their bilateral tariffs to zero (i.e. they requested FTAs with the US).

Since NAFTA enlargement was closed by a reluctant US Congress, Latin American barriers against US investment were the next dominos to fall. The US diverted FTA-requests to the Enterprise for the Americas Initiative and EAI became a swinging success. Twenty-six nations signed Framework Agreements in 1991 (these swap unilateral investment liberalisation for the promise of closer US trade relations leading eventually to an FTA). More recently, ongoing implementation of NAFTA preferences have threatened Caribbean sugar and

---

[19] The origins of the FTA are difficult to trace (the further upstream one goes, the harder it is to distinguish the main channel from the tributaries). I would, however, cite Mexico's unilateral reforms and/or US aggressive unilateralism as crucial causes.

© Blackwell Publishers Ltd 1997

clothing exporters. In reaction, Jamaica and Trinidad and Tobago — nations that were previously happy to maintain protection against US exports — applied for NAFTA membership in 1995. The trend may continue. In the words of Commonwealth Secretariat (1995):

> If the Caribbean countries do not become members of NAFTA, they will be denied tariff and non-tariff free access to the regional bloc as it expands and deepens.[20]

According to former US and Mexican trade ministers Carla Hills and Jaime Serra, *inter alios*, (see Serra et al., 1997), NAFTA also stimulated a series of overlapping bilateral and plurilateral agreements in the Hemisphere. Most important, Mercosur was established and it stimulated intra-bloc trade more than external trade. Although it covers a relatively small market (and so generates weak forces for inclusion), Mercosur has created its own domino effects. Chile and then Bolivia — nations that previously found it optimal to stay out — joined Mercosur in 1996. EU-Mercosur barriers may fall next since the two groups have initiated talks aimed at redressing the discriminatory effects of Mercosur. If the EU-Mercosur domino falls, or even seems likely to fall, US-Mercosur barriers could be next. US President Clinton, for instance, argued that the US needs 'fast-track' authority since European leaders asked South American nations 'to give Europe considerations and breaks in opening their markets and leaving us out' (*Financial Times*, 25 September, 1997).

c. European Dominos

The idiosyncratic event in Europe was the completion of the Single Market. As Section 2 describes, the Single Market initiative triggered the EEA talks and these raised the difficulties of extending single-market access to nations not participating in EU decision-making. This was soon clear to many EFTAns, but long traditions of self- or Soviet-imposed neutrality ruled out the natural solution: EU membership. A second idiosyncratic event — the weakening of the USSR — eliminated this and Austria applied in 1989. Ongoing EC92 implementation and the pending enlargement made the potential loss of competitiveness even more threatening. Each EFTA nation individually faced the prospect of losing-out in the EU-12 markets *and* in the markets of joining EFTAns. Since the combined EU-EFTA market on average accounted for three-quarters of EFTA exports, the pressure on the hold-outs mounted. Sweden put in its application in 1991, with Finland, Norway and Switzerland following in 1992. In 1995 the EU-12 became the EU-15, rather than the EU-17, since the fiercely independent Norwegians and Swiss rejected membership by referendum.[21]

---

[20] From an article in the *Financial Times*, 21 June, 1995, p. 7.

[21] Swiss rejection came with their 1992 refusal of EEA. Baldwin (1994) argues that domino effects also operated in Europe's first (1950–1960) and second (1961–1974) stages of integration.

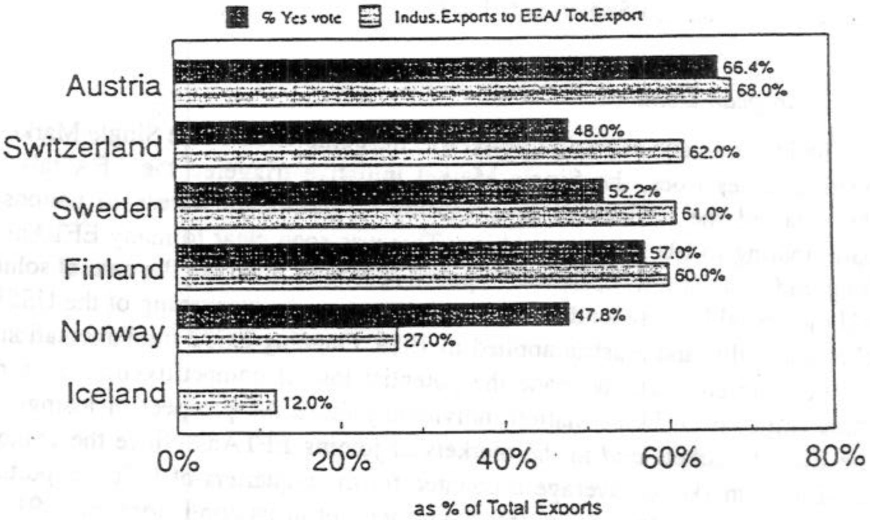
© Blackwell Publishers Ltd 1997

d. Some Political Economy Evidence

Some indirect evidence for domino effects can be found in the ordering of the four referendums on EU membership. Leaders of Austria, Finland, Sweden and Norway agreed to sequence the national votes in order of descending of EU popularity on the belief that 'Yes' votes would influence subsequent polls (admittedly, the monkey-see-monkey-do phenomenon could explain the leaders' reasoning as well). Parts of Austria were held by Soviet forces until 1955, so Austrians viewed membership as a historic opportunity to prevent such events in the future. Also, Austria is the EFTA country most heavily dependent on the EU market for its manufactured exports. For both reasons, membership was popular in Austria, so Austria would go first. Finland held its vote next. The USSR took parts of Finland in WWII and imposed 'Finlandisation', so Finnish membership also involved significant non-economic gains. Sweden came after Finland and before Norway. It was hoped that the widely anticipated positive results in Austria and Finland would tilt the fine balance in Sweden towards the 'Yes' side. Norway, a nation that fiercely guards its independence (won only in 1905), was to go last. Oil is not subject to pressures-for-inclusion and it dominates Norway's exports. Nevertheless, almost a third of Norwegian manufactured exports go to Sweden and Finland, so some domino effect was hoped for.

As Figure 1 shows, there is positive correlation between the Yes-vote share and the importance of manufactured exports to the EU plus EFTA area. The exception is Switzerland whose voters rejected even the EEA although 60 per cent of its exports face discrimination from EC92. Iceland is also included in the figure. Of all the EFTAns (leaving Liechtenstein's 26,000 citizens apart) Iceland

FIGURE 1
Importance of Manufactured Trade with EU and EFTA and Referendum Results



© Blackwell Publishers Ltd 1997

was the only one not to negotiate membership — presumably because the Icelandic anti-membership forces outweigh the pro-membership forces. A glance at the figure shows that the powerful export interests — so important in other EFTAns — are largely absent in Iceland. To put it sharply, the pro-membership forces were so weak in Iceland that a referendum was not even held. In this sense, the 'Yes' vote would have been certainly less than 50 per cent, and probably less than Norway's 48 per cent.

e. Some Econometric Evidence

Sapir (1997) uses year-by-year estimates of the standard gravity equation to test the domino theory in Europe. The specific question is whether increased integration in the EC/EU had a negative impact on non-members — especially the EFTAns — and, thereby, prompted their applications for deeper integration. According to his results, from 1960–1972 intra-EFTA trade fared relatively well thanks to preferences matching those enjoyed by EC members on intra-EC trade. EFTA exports to the EEC, however, suffered during this period due to discrimination. When the UK, Denmark and Ireland joined in the EEC in 1973, Sapir finds that from 1974 to 1985, EFTAns' exports did not suffer greatly from non-membership. From the mid-1980s, however, he finds that non-membership impinged on EFTAns' exports to the EC. He uses this evidence to support for the domino interpretation of the EEA agreement and the 1995 enlargement.

5. SUMMARY AND CONCLUDING REMARKS

This paper argues that resurgent regionalism was not caused by frustrations with the GATT negotiation process, by the complexity of modern trade barriers, or even by the US's conversion from devoted multilateralist to ardent regionalist. GATT Rounds have always been perceived as long, slow, and difficult. Moreover, old-fashioned multilateralism — in the form of the Uruguay Round — was more successful at liberalising complex barriers than the new regional arrangements (West Europe excluded). Finally, the purported role of the US's purported conversion is based on two incorrect readings of history.

The US was not multilaterism's white-knight before frustration with the GATT-process in 1982 led it to embrace the Dark Side (regionalism). The US has long been open to regional deals, negotiating bilaterals with Canada in 1948, 1958 and 1965. In 1974, the US congress authorised FTA talks with Canada, and then, in 1979, mandated a study of North American preferential trade. And all of this happened before the ink was dry on the Tokyo Round deal. In 1981, the USTR's office delivered a pro-FTA study to Congress. There may be a way to explain how sluggish progress towards a new round in 1982, or later crises in the

Uruguay Round prompted the US Congress (in 1974) and USTR (in 1981) to favour preferential trade. I, however, cannot imagine one.

It is also a mis-reading of history to view Canada and Mexico as eager participants just waiting for the US to come around to regionalism. The facts are that Canada initiated the US-Canada FTA, and Mexico initiated the US-Mexico FTA, which later became NAFTA at the Canadians' insistence. And it was NAFTA — not the US's supposedly new-found love for regionalism — that stimulated the series of overlapping bilateral and plurilateral agreements in the Americas. The important conversion in the 1980s and 1990s was the reversal of Canada's and Mexico's long-held aversion towards unrestricted imports and investment from their hegemonic neighbour. If any conversion by the US is relevant, it is the US's adoption of aggressive unilateralism, which encouraged its neighbours to seek shelter under bilateral trade accords. Finally, it should be obvious to all well-informed observers that Europe's recent regionalism (say, since 1986) had nothing to do with US attitudes towards preferential trade.

In short, the US's purposed conversion does not explain the resurgence of regionalism in Europe and the Americas. It does not, therefore, explain the incidents of regionalism that could pose a threat to the global trading system. US attitudes may have mattered in Asia, but discriminatory integration has not caught on there and is unlikely to do so (see Section 2).

Rejecting the standard explanation is important since if it is wrong, many fears concerning regionalism are misplaced. Regional liberalisation did not occur because regionalism was easier, so regional arrangements are not necessarily substitutes for multilateral liberalisation. Moreover, since the US was always prepared to pursue regional and global liberalisation in tandem, resurgent regionalism does not mean that the US is disaffected with the WTO-centered system.

This paper also seeks to explain the current wave of regionalism with a domino theory. That is, resurgent regionalism was caused by idiosyncratic events that were multiplied many times over by a domino effect. In particular, announcement of the US-Mexico FTA created powerful forces for inclusion that led to NAFTA and a sequence of NAFTA membership requests. Since these requests were put off, the forces vented themselves in a series of overlapping bilateral and plurilateral agreements in the Americas. One of these — Mercosur — is generating its own domino effects. In Europe, the triggering events were EC92 and the Soviets' fall.

The domino theory cannot explain all incidents of regionalism, since some agreements have little to do with economics. For instance, the theory never tells us why the triggering incidents of regionalism occurred. The theory's merits are that it fits the facts, and that it rests on the analysis of systematic political economy forces rather than on the opinions of a handful of policy makers in a single country.

© Blackwell Publishers Ltd 1997

a. The Consequences of Regionalism?

If the domino explanation is correct, regionalism may be a powerful force for multilateral liberalisation. To see this, a model of why nations liberalise is needed. Baldwin (1994, Chapter 3) informally proposes one such model that focuses on exporters as the key anti-protection force and import-competitors as the key pro-protection force. In this framework any liberalisation, which raises imports and exports, systematically strengthens the power of pro-trade forces (exporters) and weakens that of anti-trade forces (import competitors). To illustrate the approach, consider how episodic GATT rounds started a liberalisation-dynamic in industrialised nations. Announcement of a reciprocal trade talk multiplies the ranks of pro-trade forces by making exporters — normally indifferent to domestic protection — opponents of domestic import barriers. The same goes on in other nations, so the resulting political equilibrium involves liberalised domestic and foreign markets. While the liberalisations are phased in, export interests get stronger as they expand output and employment. Import-competitors get weaker as they scale back or shutdown. A few years down the road another GATT Round is launched and the cycle is repeated. This time with lower initial import barriers and relatively stronger pro-trade forces.

At first glance, this seems to support Larry Summers' argument that all the trade liberalisation 'laterals' (uni-, bi-, pluri- and multi-) are good. This is largely true, however, there are exceptions. For example, the EU's Common Agriculture Policy (which arose from Europe's regional liberalisation of food trade) created a powerful anti-liberalisation force. Likewise, because developing nations often have high tariffs, South-South FTAs may hinder multilateral liberalisation by creating or supporting uncompetitive, import-substituting industries.

Seen from this perspective, North-South preferential arrangements may be an excellent way of promoting liberalisation in developing nations. The point is that developing countries were exempted from the GATT-Round-dynamo, since rich nations decided to let poor nations free-ride on their tariff-cutting. Because of this notion, called 'special and differential' treatment, developing-country exporters were not required to face-down protectionists in their own countries as a condition for winning expanded market access overseas. It is hardly surprising, therefore, that GATT Rounds had such little impact on developing nations' import barriers. This is where North-South FTAs come in.

To be consistent with the GATT's article 24, FTAs require all members to bring their tariffs to zero on substantially all trade; special and differential treatment is not normally allowed. Since northern industrial tariffs are negligible (except in a few small sectors), the northern partner's industry will be globally competitive. Thus the North-South market-opening will face the southern import-competing industries with competition that is close to global-competition levels. Inefficient, anti-liberalisation industries in the southern partner will consequently

© Blackwell Publishers Ltd 1997

shrink. By direct (market access) or indirect (general equilibrium) means, North-South FTAs will also boost the sales and power of southern export interests. All this should make southern nations less resistant to further liberalisation in multilateral talks. Lastly, if the North-South FTAs have any impact on food trade, the same wedge may help break apart northern agricultural protection.

b. The Big Question: Part Three

Does recent regionalism threaten the future of the world trading system? My best guess is that because trade is already quite free in the major trading nations, few regional liberalisations are capable of creating important anti-liberalisation forces (the exceptions are likely to be South-South FTAs). For this reason, most regional deals will weaken the key opponents of free trade (import competitors) while simultaneously strengthening its key proponents (exporters). Regional integration will, therefore, foster multilateral liberalisation and vice versa, just as it has done for the past 40 years.

If this is right, regional deals are not building blocks or stumbling blocks. Regionalism is half of the trade liberalisation 'wheel' that has been rolling towards global free trade since 1958.

REFERENCES

Anderson, K. and R. Blackhurst (1993), 'Introduction and Summary', in K. Anderson and R. Blackhurst (eds.), *Regional Integration and the Global Trading System* (London: Harvester-Wheatsheaf).

Baldwin, Richard E. (1993a), 'Asymmetric Lobbying: Why Governments Pick Losers', GIIS mineo available by e-mail (Geneva).

Baldwin, Richard E. (1993b), 'A Domino Theory of Regionalism', NBER WP 4465 (Cambridge).

Baldwin, Richard E. (1995), 'A Domino Theory of Regionalism' in R. Baldwin, P. Haarparanta and J. Kianden (eds.), *Expanding Membership of the European Union*. (Cambridge: Cambridge University Press).

Baldwin, Richard E. and A. Venables (1995), 'Regional Economic Integration', in Grossman and Rogoff (eds.), *Handbook of International Economics* (Amsterdam: North Holland).

Baldwin, Robert E. and Richard E. Baldwin (1996), 'Alternative Approaches to the Political Economy of Endogenous Trade Liberalisation', *European Economic Review*, **40**, 7775–82.

Bergsten, F. (1996), 'Competitive Liberalisation and Global Free Trade: A Vision for the Early 21st Century' (Washington DC: Institute for International Economics).

Bergsten, F. (1997), 'Open Regionalism' (Washington: Institute for International Economics).

Bhagwati, J. (1993), 'Regionalism and Multilateralism: An Overview', in K. Anderson and R. Blackhurst (eds.), *Regional Integration and the Global Trading System* (London: Harvester-Wheatsheaf).

Bhagwati, J. and A. Krueger (1995), *The Dangerous Drift to Preferential Trade Agreements* (Washington: American Enterprise Institute).

Blomstrom, M. and A. Kokko (1997), 'Regional Integration and Foreign Direct Investment', CEPR DP 1659 (London).

© Blackwell Publishers Ltd 1997

Cable, V. (1994), 'Overview', in V. Cable and D. Henderson (eds.), *Trade Blocs? The Future of Regional Integration* (London: Royal Institute of International Affairs).

CEA (1991), *Economic Report of the President* (Washington: Council of Economic Advisors).

CEPR (1992), *Is Bigger Better?: The Economics of EC Enlargement* (CEPR, London).

Commonwealth Secretariat (1995), 'Impact of Uruguay Round and Nafta on Commonwealth Caribbean Countries' (London).

De Torre, A. and M. Kelly (1992), 'Regional Trade Arrangements', IMF Occasional papers, No. 93 (Washington).

Dixit, A., G. Grossman and E. Helpman (1997), 'Common Agency and Coordination: General Theory and Application to Government Policy Making', *Journal of Political Economy*, 105, 4, 752–70.

Either, W. (1996), 'Multilateral Roads to Regionalism', CIES DP (University of Adelaide).

Fernandez, R. (1997), 'Returns to Regionalism: An Evaluation of Non-traditional Gains from RTAs', CEPR DP 1634 (London).

Galal, A. and B. Hoekman (1977), *Regional Partners in Global Markets: Limits and Possibilities in the Euro-Med Agreements* (CEPR, London).

GATT (1993), GATT/1596, Press Communiqué (Geneva, 16 September).

Grossman and Helpman (1996), 'Rent Dissipation, Free Riding and Trade Policy', *European Economic Review*, 40, 795–803.

Hillman, A. (1982), 'Declining Industries and Political-support Protectionist Motives', *American Economic Review*.

Hillman, A. (1989), *The Political Economy of Protection* (London: Harwood Academic Publishers).

Hufbauer, H., J. Schott and D. Clark (1994), *Western Hemisphere Economic Integration* (Washington: Institute for International Economics).

Krueger, A. (1993), 'Free Trade Agreements as Protectionist Devices: Rules of Origin', NBER WP 4352 (Cambridge MA).

Krugman, P. (1993), 'Regionalism versus Multilateralism: Analytic Notes', in J. De Melo and A. Panagariya (eds.), *New Dimensions in Regional Integration* (Cambridge: Cambridge University Press for CEPR).

Lawrence, R. (1996), *Regionalism, Multilateralism and Deeper Integration* (Washington: Brookings Institute).

Panagariya, A. (1996), 'APEC and the United States', CIES DP (University of Adelaide).

Spair, A. (1997), 'Domino Effects in Western European Trade, 1960–1992', CEPR DP (London).

Schott, J. (1988), 'The Free Trade Agreement: A US Assessment' in J. Schott and M. Smith (eds.), *The Canada-US Free Trade Agreement: The Global Impact* (Washington: Institute for International Economics).

Schott, J. (1989), 'More Free Trade Areas?' in J. Schott (ed.), *Free Trade Areas and US Trade Policy* (Washington: Institute for International Economics).

Schott, J. and G. Hufbauer (1992), 'NAFTA: Questions of Form and Substance', in L. Waverman (ed.), *Negotiating and Implementing a North American Free Trade Agreement* (Vancouver: The Fraser Institute).

Serra et al. (1997), *Reflections on Regionalism: Report of the Study Group on International Trade* (New York: Carnegie Endowment for International Peace).

Smith, M. (1988), 'The Free Trade Agreement in Context: A Canadian Perspective', in *The Canada-United States Free Trade Agreement: The Global Impact* (Washington: Institute for International Economics).

Summers, L. (1991), 'Regionalism and the World Trading System', in L. Summers (ed.), *Policy Implications of Trade and Currency Zones* (Federal Reserve Bank of Kansas City).

USTR (1981), *North American Trade Agreements* (Washington: Office of the US Trade Representative).

Whalley, J. (1993), 'Regional Trade Arrangements in North America: CUSTA and NAFTA', in J. De Melo and A. Panagariya (eds.), *New Dimensions in Regional Integration* (Cambridge: Cambridge University Press for CEPR).

Whalley, J. (1996), 'Why Do Countries Seek Regional Trade Agreements?', NBER WP 5552

© Blackwell Publishers Ltd 1997

(Cambridge MA).

Winters, A. (1996), 'Regionalism versus Multilateralism', World Bank Policy Research Working Paper 1687 (Washington).

Wolf, M. (1994), *The Resistable Appeal of Fortress Europe* (Washington: American Enterprise Institute).

Wonnacott, P. (1987), *The United States and Canada: The Quest for Free Trade* (Washington: Institute for International Economics).

WTO (1995), *Regionalism and the World Trading System* (Geneva: WTO Secretariat).

Yeats, S. (1997), 'Does Mercosur's Trade Performance Raise Concerns About the Effects of Regional Trade Arrangements', World Bank Policy Research WP 1729 (Washington).

Yi, S. (1996), 'Endogenous Formation of Customs Unions Under Imperfect Competition: Open Regionalism is Good', *Journal of International Economics*, **41**, 151–75.

_______ (1995b). "Communiqué of the Meeting of the OECD Council at Ministerial Level, May 1995" (Paris: OECD), mimeo.

_______ (1995c). *Introduction to the OECD Codes of Liberalisation* (Paris: OECD).

_______ (1994). *National Treatment in Established Foreign-Controlled Enterprises* (Paris: OECD).

_______ (1992a). *The OECD Declaration and Decisions on International Investment and Multinational Enterprises: 1991 Review* (Paris: OECD).

_______ (1992b). *Foreign Direct Investment Policies and Trends in the 1980s* (Paris: OECD).

Transnational Corporations, vol 4, no 2

(Au 5)

Regional integration in a globally networked economy

LIBRARY DOCUMENT DELIVERY

Stephen J. Kobrin*

Much of the discussion on regional integration is framed in terms of "building blocks" or "stumbling blocks"; will regional economic agreements, such as the European Union or the North American Free Trade Agreement, facilitate the development of an open and liberal international economy, or do they portend the devolution of the world economy into closed, competing economic areas? It is argued here that an open international economy is not an artifact of political processes or of the current distribution of political and economic powers. Rather, the emerging globally integrated world economy is a reflection of fundamental changes in the scale and complexity of technology and the digitalization of international economic transactions. A future of closed regionalism is unlikely, as even the largest economic regions are too small to be viable economically, and there is no reason to believe that regional borders will be any more meaningful than national borders in the face of an electronically networked world economy. Regions are national markets writ large at a time when geography as a basis for the organization of economic activity is in question.

> *The splitting-up of the world into three great superstates was an event which could be and indeed was foreseen before the middle of the twentieth century. With the absorption of Europe by Russia and the British Empire by the United States, two of the three existing powers, Eurasia and Oceania, were already effectively in being. The third, Eastasia, only emerged as a distinct unit after another decade of confused fighting. The frontiers between the three superstates are in some places arbitrary, and in others they fluctuate according to the fortunes of war, but in general they follow geographical lines.* (Orwell, 1949, pp. 152-153).

---

* William H. Wurster Professor of Multinational Management, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania, United States. The author would like to thank John Ikenberry and Alan Rugman for their comments on a previous draft.

WHARTON REPROGRAPHICS

43

At century's end, the world economy presents a paradox of simultaneous integration and fragmentation as both the centripetal and centrifugal forces of globalization and regionalization are evident to even a casual observer. Can they continue to coexist, or will one come to dominate? Will the late 1990s witness progress towards an integrated, open and liberal international economic order or a devolution into relatively closed regional blocks, the economic analogs of George Orwell's "three superstates"?

On the one hand, we appear to be in the midst of a transition from an international world economy rooted firmly in national markets to integrated global systems of production and technology. Internationalized production has replaced trade as the "glue" binding international economic transactions: by the early 1990s, 37,000 parent firms with worldwide sales of about $4.8 trillion controlled about one third of the world's private-sector productive assets (UNCTAD-DTCI, 1994).

During the late 1980s foreign direct investment (FDI) grew at an average annual rate nearly three times above that of trade and four times above that of world output. Critically, 1990 was the first year in which worldwide sales of foreign affiliates exceeded world exports (United States, Office of Technology Assessment, 1993, p.1). Furthermore, the exponential growth of cross-border strategic alliances or collaborative agreements during the early 1990s may well herald the emergence of a complex networked global economy (Mytelka, 1991).

At the same time, there has been a dramatic increase in regionalization. The European Union and the 1992 Single Market are a reality, as is the enlargement of the European Union to fifteen countries. Further expansion of the European Union with at least some former Central European economies becoming members around the turn of the century is a real possibility. Despite the problems resulting from the crisis in Mexico, the North American Free Trade Agreement (NAFTA) is solidly in place, and prospects for MERCOSUR appear promising. In the Asia-Pacific region, the first steps have been taken towards the removal of all trade barriers, and discussions on regional integration are taking place within at least three sub-regional groups: Asia-Pacific Economic Cooperation (APEC), the Association of South-East Asian Nations (ASEAN) and South-Asian Association for Regional Cooperation (SAARC)). It has been suggested that an Indian Ocean regional economic grouping might even be possible.[1]

---

[1] Niki Tait, "Indian Ocean grouping faces a big 'maybe' ", *Financial Times*, 14 June 1995, p. 8.

In fact, the General Agreement on Tariffs and Trade (GATT) was notified of 33 regional integration agreements between 1990 and 1994—a 300 per cent increase compared with the entire decade of the 1980s and almost 30 per cent of all international agreements since 1948.[2] Intra-regional trade (expressed as a region's share of the total external trade of its constituent markets) grew substantially in East Asia, North America and Europe during the 1980s.[3]

There is considerable difference of opinion about the resolution of this apparent paradox of regionalism versus multilateralism. Some, such as Robert Z. Lawrence (1995), square the circle by arguing that regional agreements are "building blocks" which will facilitate multilateralism and lead to an integrated and open global economy. Others see regionalization as "stumbling blocks", as a withdrawal towards closure which may lead to a reduction in inter-regional economic activity and inhibit further development of a liberal international economy. Jagdish Bagwati, for example, has argued that "This obsession with free trade areas strikes a blow at the multilateral trading system in ways that are either ill-understood or deliberately discounted, disregarded and distorted by politicians and lobbyists."[4]

The most pessimistic are those who believe that globalization has reached its limits—that the glass is half empty and draining. Paul R. Krugman (1992) argued that we are experiencing the decline of the second global economy, a devolution into competing regional blocks which will inevitably become more restrictive against trade from the outside. The obvious parallel is to the 1930s when what Charles Kindleberger has called "disarticulation"—the containment of trade within block borders—prevented a return to the integrated pre-First World War international economy and plunged the world more deeply into depression.

Two related arguments are presented here. First, the emerging globally integrated economy is not an artifact of the current distribution of political and economic powers or the outcome of a successful negotiating process, but rather reflects very basic, fundamental and secular changes in the scale and complexity of economic and technological activities. Second, regional markets are, at best, national markets writ large. There is no reason to believe a priori that they are large enough to be viable economically, or that regional borders will be any less permeable than national borders in an electronic age. A future restricted or closed regionalization is therefore not a likely scenario for the world economy.

---

[2] Guy de Jonquières, "WTO's blessing for trade groups", *Financial Times*, 27 April 1995, p. 8.

[3] Between 1980 and 1990, intra-regional trade grew from 33.8 per cent to 38.4 per cent in East Asia, 31.5 per cent to 37.3 per cent in North America and 52.4 per cent to 63.4 per cent in the European Community (Kirkpatrick, 1994).

[4] Jagdish Bagwati, "The high cost of free trade areas", *Financial Times*, 31 May 1995, p. 13.

A discussion of regionalization in the context of the increasing scale of economic activity and the digitalization of cross-border transfers is presented first. In answering the question of whether geography remains a meaningful mode for the organization of economic activity, it is suggested that the ambiguity of medieval Europe may well be an appropriate analogy for the post-modern world economy. The conclusion outlines some policy implications for governments.

A world of regions?

A regional scenario is not inconsistent with events. First, *multilateralism appears to have been pushed to its limits*. Although an agreement was reached after nine years of negotiations in the context of the Uruguay Round, the process was tortuous and the outcome far from clear until the very end. That experience did not leave most observers with a great deal of confidence in the success of another multilateral round. What confidence remained was drained further by the battle over the choice of the head of the World Trade Organization (WTO) (GATT's successor), which bogged down in disagreement among the three major regions, and the United States-Japan bilateral auto dispute that threatened to render the WTO's critical new adjudication procedures stillborn. Furthermore, substantial domestic opposition to multilateralism—and particularly its institutions—has surfaced in a number of countries, particularly the United States.

Second, *regionalism is very much a reality*. Despite all of the bumps, detours and delays on the road to achieve integration, the European Union's single market exists, albeit in somewhat incomplete form. Although NAFTA is a less ambitious effort—its objectives do not go much beyond a free trade area—it has been more successful and has presented fewer problems than even its most ardent supporters had hoped for, despite the difficulties raised by the recent Mexican crisis. Furthermore, the eventual expansion of NAFTA to include other countries (Chile is the most likely first candidate) is a distinct possibility; and with MERCOSUR for the first time in decades, economic integration is actually taking place within Latin America.

There is increasing discussion of economic links among APEC members and a formal agreement in principle on the removal of barriers to trade within the region—in twenty-five years—has been reached. Although, to date, economic integration within Asia has been driven by private-sector initiatives—primarily the vertical integration of Japanese transnational corporations (TNCs), there is discussion among South-Asian countries on moving towards the creation of a free trade area.[5]

In short, while the intensity and pace of integration differs considerably in Europe, North America and Asia, regionalization is a reality and, at least in the short run, the future of regional blocks appears consistent with the progress of events.

Third, *regionalization may be more consistent with the structural limitations of the international system than an integrated global economy*. The argument is familiar: the establishment and maintenance of an open, liberal, global economy may require a dominant power or *hegemon* to set the rules and bear the costs of supplying the public goods requisite to the free flows of trade and investment. Given the end of the Cold War, when security issues often overrode economic concerns, and the decline in the *relative* dominance of the United States, an open global economy requires cooperation in a multipolar world comprised of a number of relatively equal powers. That is a difficult task at best. To be clear, cooperation in the absence of a dominant power is not impossible; indeed, there is a large literature in international relations that argues that this is feasible under a number of very plausible conditions. But it cannot be achieved easily, and regionalization is certainly as—if not more—consistent with what theory postulates.[6]

Building blocks or stumbling blocks?

Will regionalization lead to integration or devolution? Optimists such as Robert Z. Lawrence (1995) argued that regional blocks will be "building blocks'' paving the way to an open global economy. They note that we may well have reached the limits of multilateralism given the complexity of the issues, the number of actors involved (there were 108 participants in the Uruguay Round), the differences between them and the lack of clear international economic leadership.

---

[5] The Asia-Pacific Economic Cooperation (APEC) organization was created in 1989 as a forum for regular discussions on regional trade issues and economic cooperation. Its membership comprises 17 economies: Australia, Brunei Darussalam, Canada, China, Hong Kong, Indonesia, Japan, Republic of Korea, Malaysia, Mexico, New Zealand, Papua New Guinea, Philippines, Singapore, Taiwan Province of China, Thailand and the United States. In mid-November 1994, the APEC ministers endorsed a set of non-binding investment principles.

[6] The debate about whether cooperation is probable or even possible in the absence of a dominant or hegemonic power plays a central role in the literature of International Political Economy. The idea that a dominant power is necessary to provide order and the "public goods" necessary for the maintenance of an open, liberal international system originated with Charles Kindleberger's work on the Great Depression; for a recent summary of this literature, see Baldwin (1993).

As the argument goes, it should be easier to reach agreements maintaining free flows of trade and investment within regions where the issues are less complex, substantial economic exchanges already exist, the number of actors is reduced, and geographic—and perhaps cultural—proximity facilitates the negotiating process. Regionalization or ''minilateralism'' would promote free flows of trade and investment within blocks and, in time, lead to multilateral negotiations between a reduced number of participants (the blocks negotiating as entities) and an open, liberal global economy.

The critical assumptions here are that regional blocks will be trade creating (rather than trade diverting)[7] and would maintain relatively open borders. That is far from assured; the motivation for regional integration is not an overall belief in a liberal world order, but a conviction that regionalization is good for the countries involved. There appears to be no more reason to assume that the borders of regional blocks will be relatively open than to assume they will be relatively closed. ''Fortress Europe, NAFTA or Asia'' may be as plausible a scenario as ''building blocks''. Paul R. Krugman (1992, pp. 12-13) argued strongly that the idea of relatively open blocks is nonsense; that block insiders will always have a stronger voice than outsiders; and that the result will be a world of regions which ''will become more restrictive against trade from the outside''.

Relatively closed regional blocks would have differential impacts on trade and FDI depending on the nature of the commodity or the strategy of firms. Transnational corporations in industries such as processed foods or consumer products in which inter-regional trade in intermediate or finished goods is very limited might suffer only minor disadvantages, or even gain, from ''fortress'' regionalism. Their traditional strategic focus has been the foreign affiliate, and many are only now in the process of a transition to regional strategies and organizational structures. However, TNCs in industries that are relatively integrated globally, in which substantial inter-regional exchanges of goods and technologies are the norm could suffer serious consequences from closed regionalization. Firms in aerospace, electronics, semiconductors or pharmaceuticals in which the scale of technology requires inter-regional markets to afford a competitive research-and-development budget would find closed regionalization catastrophic.

A forecast of devolution into regional blocks assumes that the degree of multilateral cooperation necessary to sustain a relatively open global economy is not possible in a multi-actor, multipolar world. Regional integration represents a compromise: an area small enough for multilateral negotiations to be tractable, but large enough to provide for gains from trade and scale economies. Whether regionalization is taken as a positive or negative development *vis-à-vis* an open global economy depends on the assumptions about the motivation for, and character of, the blocks.

Regionalization is not the answer

Much of the current discussion on regionalization is framed either explicitly or implicitly in a context more appropriate to the 1930s than the 1990s. The motivation for the current attempts towards regional integration differs considerably from that which drove ''disarticulation'' during the Great Depression. In fact, the very term *block* entails atavistic assumptions. Although protectionist sentiments still abound, the primary motivations for present-day integration are efficiency and scale rather than containment and protectionism.

Regional economic integration in the 1990s should not be seen in isolation; it is the latest step in a century-long process of geographic expansion of markets from local, to national and now to international. In fact, a global world economy has been defined in terms of the cross-border integration of national economic spaces (UNCTAD-DTCI, 1994). The expansion of markets in geographic space has been driven by efficiency considerations. The process began in earnest in the mid-nineteenth century with the revolution in manufacturing technologies and the application of science and technology to industry. Larger markets allowed gains from specialization (division of labour), differences in resource endowments, and from scale economies in manufacturing and technology. There is no reason whatsoever to assume that consolidation of markets at the regional level represents the final step of this process.

Those who argue that the world economy is devolving into three relatively closed regional blocks make the following assumptions:

- Regional markets are large enough to be viable economically.

---

[7] For a good discussion of this issue, see *The Economist*, ''Building blocks or stumbling blocks'', 31 October 1992, p. 69.

they do not resolve the emerging asymmetry between economic and political geography; they are a linear solution in a world of increasing complexity and ambiguity.

There is no reason why a larger market area should make discrete borders more meaningful or territorial control of economic activity more effective in an electronically networked world economy. Regionalization is a modern solution to a post-modern problem. Moats and drawbridges did not provide a meaningful defence against the cannon, and regional borders will not inhibit cellular transmissions of data and information. Economic activity can no longer be contained geographically at the country or regional levels.

Back to the future?

Almost twenty years ago, Hedley Bull (1977) argued that the emergence of a modern and secular counterpart of Western Christendom, with its characteristic overlapping authority and multiple loyalties, was within the realm of possibility. The post-modern future may well resemble more closely the medieval past—at least metaphorically—than the more immediate, geographically organized world of national markets and nation-States to which regional blocks clearly belong. Although medieval ''states'' occupied geographic space, politics were not organized in terms of unambiguous geography. Political authority took the form of hierarchical, personal relationships of often overlapping and intertwined mutual obligations and rights. Borders were diffuse, representing a momentary projection of power rather than the limit of sovereignty (Ruggie, 1993).

While the medieval analogy has very obvious limits, the past may well contain applicable lessons for the future. A neat, unambiguous, mutually exclusive geographic ordering of economic and political authorities may no longer be the norm. Borders are diffuse and permeable, compromised by transnational integration and global telecommunications. Relationships are increasingly networked rather than hierarchical, with both individuals and organizations enmeshed in complex worldwide webs. Multiple and competing loyalties result.

Both public policy makers and private-sector managers must deal with competing systems for the organization of economic and political authority: sovereign and non-sovereign actors; geographic and electronic markets; local politics and global economies; fragmenting States and regional blocks; and bilateralism and multilateralism. States experience increasing constraints on internal autonomy and find territoriality a problematic basis for exercising economic and political control. Transnational corporations find themselves caught between conflicting jurisdictions, each of which believes it has the right to exercise economic authority.

A world in transition

Much of the discussion of globalism versus regionalism implies an assumption of cyclicality: the ebb and flow of integration and devolution of the world economy. As ''disarticulation'' of the first globally integrated economy was a cause for the Great Depression, the breakdown of the world economy into regional blocks portends the decline of the second. Is it possible, however, that the uncertainty and instability now being experienced in the international economy may be a reflection of secular rather than cyclical changes; of the initial stages of a systemic transformation of the organization of economic, and perhaps political, authority—a transformation perhaps comparable to the transition from a medieval to a modern state system in Europe during the sixteenth and seventeenth centuries.

In many industries, economic and technical systems are inherently global and transcending geography, being reorganized as worldwide electronic networks. On the other hand, politics remain geographic and local, firmly rooted in sovereign, territorial-based nation-States. For the foreseeable future the world will be torn by an asymmetry between the scope and mode of economic and political structures, by ambiguity of jurisdiction and by conflicting pulls on loyalty. That tension was reflected clearly in much of the debate in the United States surrounding both NAFTA and the ratification of the new GATT agreement. The main concern was the potential loss of local control and United States sovereignty, a concern that reflected the fear that national values and preferences in areas as diverse as workers' rights, the environment and lifestyles will become meaningless in a world in which sovereignty is transferred to some distant, ill-defined regional or international authority. In industries such as pharmaceuticals, aerospace and semiconductors, however, further integration or devolution of the global economy is not in the hands of the United States Congress, the European Parliament or the Japanese Diet. A global world economy is not dependent on the particular constellation of political forces at the moment. The emergence of an integrated global economy during the past decade reflects basic structural change: dramatic increases in the scale of technology, global electronic information systems and the corresponding development of network forms of organization.

however, are but one of a number of historical modes of organizing economic and political authority, and a relatively recent one at that. It is reasonable to ask whether the very idea of a geographically organized market will continue to be relevant in the emerging post-modern, electronically networked world of the twenty-first century.

The international financial world provides an example and, perhaps, a metaphor. If one is asked to name the three international financial centres, the immediate reply is London, Tokyo and New York. No one would say the United Kingdom, Japan or the United States. The reason is obvious: in this context, the three countries are not relevant as geographically based centres of political and economic authority. The international financial world is organized electronically rather than geographically; it consists of hundreds of thousands of computer screens located worldwide. Communication among them is more rapid than in a small village market: information is shared instantly across the globe. When transactions, of a magnitude not even imaginable a quarter of a century ago, consist of electronic impulses transmitted at the speed of light, geography and borders lose meaning as organizational constructs.

In the post-war period, TNCs replaced trade as the primary means of organizing international economic transactions; production has become internationalized. In more technical terms, firms' administrative hierarchies have replaced the market as the mode of international economic organization. The increased importance of collaborative agreements or strategic alliances, especially in high-technology industries, compared with equity-based FDI by TNCs, represents a further evolution in the primary mode of international economic organization. Cross-border strategic alliances represent the internationalization of network forms of economic organizations.

Many large international firms now are enmeshed in complex webs of multiple and polygamous alliances and cooperative agreements (Dicken, 1994). In a number of industries—automobile assembly is an excellent example—large vertically integrated chains of production have been disassembled and put back together as networks linking parts of organizations; a business unit or plant with a research unit or a supplier. These networked links, which are often cross-border, are completely dependent on integration though electronic information systems and satellite telecommunications.

Does the geographic organization of economic authority still retain any meaning either nationally or regionally? Do the concepts of home and host country apply to an IBM-Toshiba-Siemens alliance for developing the next generation of semiconductors? If it is successful who gets to control the technology? If the Boeing-Airbus collaboration to develop a super jumbo jet were to be successful and if, as it is claimed, the world market really can only support one manufacturer, which jurisdiction should have responsibility for the control of the resulting global natural monopoly?[13] Again the same question must be asked: why should geography remain viable as a mode of economic organization simply because the market area has increased from national to regional?

At century's end we face an asymmetry between the modes of economic and political organization. Economic systems are increasingly global in scope and electronically, rather than geographically, organized. (To be clear, it is not argued that national markets have lost all meaning; rather, they are no longer the basic building blocks of the world economy.) Political authority, on the other hand, is local and organized geographically in terms of territorial nation-States. The viability of regional integration, and the likelihood of the devolution of the world economy into rather closed regional blocks, must be considered in that context.

Regional blocks are not the solution

Regional integration represents the extension of national economic and, perhaps, political organizations and structures over a larger geographical area. There is no question that regional integration can provide benefits to the countries involved in terms of specialization (gains from trade) and scale and scope economies. The North American Free Trade Agreement encourages a more efficient production on a continent-wide basis, and the European Union allows for scale economies both in manufacturing and research and development. The questions at hand, however, are whether or not the region represents the limit of an open, liberal economic system and whether or not the global economy is devolving into three relatively closed regional blocks. The answers to both questions are no. First, in many strategic industries the immense costs and risks of technology development, and particularly its rapidly increasing complexity, mandate cross-regional integration. Second, regional blocks represent an extension of the modern form of territorially based economic structures in geographic space. Although regional blocks may represent the limit of our ability at present to cooperate internationally, underway with a number of Central European economies. There is a real possibility of expanding NAFTA to include Chile (and later other Latin American countries). The Asia-Pacific region is very broadly defined to include virtually every country touching the Pacific Rim.

---

[13] At this point, the collaboration appears to be on "hold" due to disputes over the viability of the market.

The emerging integrated global economy is not an artifact of political relationships. It is dependent neither on the way political and economic powers are distributed among States nor on any successful outcome of negotiations designed to enhance economic efficiency through multilateral co-operation. At least in the longer run, it is not conditional on any given political-economic process. The integrated global economy is a manifestation of fundamental and irreversible changes in industry structure; it reflects the scale (cost and risk) and complexity of technologies in strategic industries (semiconductors, aerospace, telecommunications and pharmaceuticals). These industries *cannot* be organized on a uninational basis and remain technologically competitive at the same time. It can be argued that even intra-regional organization is problematic. These industries are inherently global in nature.

Putting aside blatant technological determinism (i.e., firms in strategic industries either collaborate transnationally or forego competitive development), the question here—which is certainly debatable—is whether regional blocks such as the European Union, the Americas or Asia-Pacific represent markets of sufficient diversity and size to allow firms to deal with the scale and complexity of technologies within their borders. Looking at the existing pattern of strategic alliances in electronics, semiconductors, aerospace, etc., one would think not. In this context, "Fortress Europe" may mean no more than the "walling in" of obsolete technology.

Can regional borders be closed?

In a world of tangible flows in which raw, intermediate and finished goods are traded through arm's-length transactions and in which currencies are physically transported from one market to another, controlling cross-border transactions is difficult, but possible. Borders are less meaningful as economic constructs when production is international; when affiliate sales are higher than exports; and a large proportion of existing flows are intra-firm rather than arm's-length. One has to ask if borders will have any meaning in a world in which the most important international flows are knowledge transmitted instantaneously across worldwide electronic webs. A world in which, in Nicholas Negroponte's (1995) words, "trade in atoms" has been replaced by "trade in bits".

Borders are almost entirely irrelevant in international financial markets where $1 trillion worth of foreign exchange is traded on an average day. There are severe limits on any government's ability to "defend its currency" in the face of huge flows of electronic transactions which are consummated in the blink of an eye regardless of the distance involved.[11] In industries in which the critical flows of knowledge, technology, information or capital take place electronically, national borders are rapidly losing meaning as economic constructs or as frontiers across which economically valuable flows can be monitored and controlled. The problematic nature of borders will only be exacerbated as the world moves into the age of wireless data communications: cellular modems linked worldwide through satellite networks such as Iridium, Globalstar or Immarsat.

A world of closed regional blocks with limited exchanges between them assumes that cross-border transactions can be controlled. The critical question here is why should borders around a region be more meaningful (or less permeable) than those surrounding a national market? Does it matter to a satellite whether an impulse is transmitted from New York to Minneapolis or to Melbourne? If anything, regional borders, comprising a series of national borders that are joined imperfectly, could be compromised more easily.

The increasing permeability of borders is symptomatic of a more general question. Both the modern State system and the very idea of a market—whether local, national or regional—are inherently territorial constructs.[12] One must ask whether the modern, geographically based modes of economic and political organizations are still viable in the face of emerging global electronic networks.

Is geography still relevant?

Unambiguous, mutually exclusive territorial jurisdictions separated by discrete and meaningful borders have been characteristic of the modern era from the mid-seventeenth to the late twentieth century; indeed, territoriality defines political modernity (Ruggie, 1993). Regional integration, even at its extreme of political union, merely extends territorially based modes of economic and political organizations in geographic space—albeit not always contiguously. Geographically defined national markets and nation-States,

---

[11] For a recent discussion of the problems of regulation in an electronic global economy, see Herring and Litan (1995).

[12] See Ruggie (1993) and Cooper (1986) for further discussions on this point.

- Regional borders are meaningful and economic transactions across them can be controlled.
- Economic activity can and should be organized in terms of discrete geography.

It is argued here that these assumptions are becoming questionable. First, even regional markets may not be large enough to support efficient production or competitive research-and-development efforts. Second, borders are losing meaning in an electronic age. And last, geography is becoming less relevant as *the* basis for organizing and controlling economic activity. Why should the borders of geographically defined regional markets be any less problematic than the borders of national markets in the face of an electronically networked global economy?

Regional markets are too small

A large number of strategic industries is characterized by rapidly escalating research-and-development costs combined with shortened product life cycles. For example, development costs for semiconductors increased tenfold, from $100 million for a one-megabyte chip in 1985 to an estimated $1 billion for a 256-megabyte chip in 1999. And the cost of bringing a single new prescription drug to the market is now about $300 million.[8] In these industries, technological development cannot be sustained by any single national market. Firms need to have a global market—or at least be TNCs—in order to afford a competitive research-and-development budget. Furthermore, it is far from clear that even the largest global firms can continue to absorb the costs and risks of technological development on their own. There has been an exponential increase in strategic alliances (international collaborative agreements) between TNCs from the major industrial countries during the past decade, and many of these are clearly technologically driven. According to one estimate, high-technology strategic alliances' compound annual growth rate over the 1980s was 31 per cent.[9] International collaboration is becoming the norm even in defence-related industries in which national security has traditionally been seen as far more important than efficiency, and military capability has taken precedence over economics.

---

[8] See Daniel Green, "Takeover fever: the wave of acquisitions and mergers in the drugs sector is set to continue", *Financial Times*, 22 August 1995, p. 10.

[9] For data on the growth of alliances, see Gomes-Casseres (1993) and Mytelka (1991). The growth figure was cited by Lewis Platt, Chief Executive Officer of Hewlett Packard at the Wharton School, University of Pennsylvania, on 12 September 1993.

The costs and risks of technology are one reason for the emergence of strategic alliances. A recent analysis of over 4,000 strategic alliances found that the primary motive for cooperation is the need to cope with the complexity and interrelatedness of different fields of technology (Hagedoorn, 1993). The range of technologies that must be brought to bear on product and/or process development is increasingly beyond the core competencies of even the largest and most global firms. Technologies have become so complex and change so rapidly that even industry leaders cannot master them internally: cross-firm collaboration is mandated.

If the largest national markets are too small to support competitive development efforts given the costs, risks and complexities of technologies in strategic industries, are regional blocks large and diverse enough to accomplish these objectives? It is argued here that they are not, and that the critical reason is probably complexity rather than cost and risk considerations. Both, however, are relevant and important.

While the three major regions are not by any means homogeneous, North America, Europe and East Asia have developed different approaches to capitalism and have different strengths when it comes to research, development and the introduction of new products. Harnessing these interregional differences may well provide a competitive advantage given the complexity of technology and the fast pace of its development. While hard data on alliances are still difficult to come by, there is a general agreement that almost all involve companies from the United States, Europe and Japan, and that close to half entail inter-block (i.e., United States-European) arrangements (United States, Office of Technology Assessment, 1993, p. 119). An executive of a major Japanese pharmaceutical company argued recently that the advantages of inter-regional collaboration are a significant motivation for maintaining research laboratories in Japan, London and Boston; his company must utilize the different approaches to problem solving of professional staff in each region to succeed.[10] Similarly, a study of semi-conductors and biotechnology (Kogut, Walker, Shan and Kim, 1994) concluded that the inter-regional distribution of technological capabilities is an important determinant of emerging global networks.

In fact, "outward pressure" exists in all regional markets. In the European Union, widening appears to have taken precedence over deepening; Austria and two Nordic countries have joined and active negotiations are

---

[10] Speech by Yamamoto of Eisai Co., Ltd at the "Strategic Management Society" annual meeting at HEC, Jouy-en-Josas, France, 22 September 1994.

In many strategic industries globalization presents a discrete rather than a marginal calculus. The issue is not one of a choice between autonomy and increased efficiency or gains from trade; of a trade-off between local production and the affordability of goods. Unless it is possible to call a halt to technological innovation—to forego the next generation of memory chips, pharmaceuticals, global cellular communications or super jumbo jets—transnational integration, increasingly taking the form of transregional alliances, is here to stay. The question is how it can be effectively dealt with and what it means in terms of "national competitiveness" and the ability to achieve national goals.

Current trends towards regionalization must be seen in this context. To ask whether regional blocks are "stumbling blocks" or "building blocks" assumes that global integration is a political process and that it depends on the effectiveness of multilateral negotiations. Although the character and depth of integration differs dramatically across all three major regions, each represents an attempt to extend markets geographically to exploit the gains from trade and deal with the scale of technology. While drawing a border around a region always implies the possibility of exclusion, protection is not the primary motive of the European Union, NAFTA or APEC. Regional integration is neither a building nor a stumbling block; it is a process taking place within the larger context of an integrated and transregional world economy.

To argue that global integration is a technical or economic rather than a political process is not to argue that all roads leading to Rome (another neo-medieval metaphor?) are equidistant, equally efficient, or equally comfortable. There are a number of policy implications for easing the transition:

- ***Encourage "open" regional agreements.*** Regionalization should facilitate free flows of trade and investment within regions without diverting existing external trade. Local content regulations, for example, should reflect a balance between promotion of intra-regional integration and the maintenance of global sourcing arrangements. Regionalization should enhance, rather than detract from, the trend towards liberalization of conditions for entry (and operation) of FDI now present in most developed and developing countries.

- ***Encourage "widening" rather than "deepening".*** Regional integration should aim at extending regional agreements to include all countries which will "play by the rules" rather than at the deepening of existing agreements, to encompass monetary integration, for example.

While this is not the place to debate issues such as the single European currency, the "deepening" of a wide variety of transnational linkages of production, finance and technology is taking place daily. At this point, it is more important that existing regional blocks be extended as widely as possible to obtain and maintain commitment to an open economy than it is to try to create larger ersatz "national" markets through the creation of deeper bloc-wide institutions. This should be taken as a statement of priorities and not as an argument against further European economic and political integration.

- ***Regional integration must deal with both investment and trade.*** As noted above, in 1990 the value of foreign affiliate sales was higher than that of world trade. In fact, it is increasingly difficult to separate trade and investment: they are two sides of the same coin. Regional agreements must facilitate free flows of FDI and trade as well as facilitate the cross-border integration of production and technology. Again, at this point what is required is the elimination of barriers rather than the creation of new institutions; efforts such as the recent OECD negotiations on multilateral investment rules should be encouraged (Witherell, 1995).

- ***Harmonize national and regional legislation and encourage multilateral organizations.*** Problems resulting from the asymmetry of economic and political organizations and structures will increase rapidly in the future; dealing with the anti-trust implications of global national monopolies is but one example. As networked alliances increase in importance, mechanisms for dealing with both trade and FDI may become irrelevant. While the resolution of national differences through international organizations raises difficult issues, experiences with standard setting indicate that the harmonization of national (and regional) legislation is possible. The experience of the European Union, for example, shows that harmonization can be a means to effect integration and, at the multilateral level, it may be the only possible medium-term solution.

- ***Develop formal mechanisms to encourage inter-regional cooperation.*** The United States plays a pivotal role in this regard as it is a member of both NAFTA, APEC and the Atlantic community. It is important that inter-regional agreements be encouraged and undertaken as a step towards multilateralism rather than the marshaling of forces to increase the competitiveness of existing blocks. The formation of institutionalized mechanisms to insure inter-regional cooperation and communication should be encouraged, perhaps through an interregional "summit" accompanied by ongoing interaction at the ministerial level. Thus, the broad definition of APEC to encompass virtually every country touching on the Pacific rim should be encouraged. Similarly, increased trans-Atlantic economic integration should be seen as a positive step as long as it is inclusive rather than exclusive and in accord with multilateralism and the World Trade Organization.

The current round of regionalization represents both the evolutionary extension of the scope of markets across national borders and, perhaps, a final attempt to organize markets (and technology) geographically. In the longer run, how regionalization is dealt with will not affect the emergence of an integrated global economy. It will affect the distribution of benefits and costs resulting from that transition, as well as the "competitive" position of individual economies and the welfare of individual countries. ■

References

Baldwin, David A., ed. (1993). *Neorealism and Neoliberalism: The Contemporary Debate* (New York: Columbia University Press).

Bull, Hedley (1977). *The Anarchical Society: A Study of Order in World Politics* (New York: Columbia University Press).

Cooper, Richard N. (1986). *Economic Policy in an Interdependent World* (Cambridge: The MIT Press).

Dicken, Peter (1994). "The Roepke lecture in economic geography. Global-local tensions: firms and states in the global space-economy", *Economic Geography*, 70 (April), pp. 101-102.

Gomes-Cassares, Benjamin (1993). "Computers, alliances and industry evolution", in David B. Yoffee, ed., *Beyond Free Trade: Firms and Governments in Global Competition* (Boston: Harvard Business School Press), pp. 79-128.

Hagedoorn, John (1993). "Understanding the role of strategic technology partnering: interorganizational modes of cooperation and sectoral differences", *Strategic Management Journal*, 14, 5 (July), pp. 371-385.

Herring, Richard J. and Robert E. Litan (1995). *Financial Regulation in the Global Economy* (Washington, D.C.: The Brookings Institution).

Kirkpatrick, Colin (1994). "Regionalization, regionalism and East Asian economic cooperation", *World Economy* (March), pp. 191-202.

Kogut, Bruce, Gordon Walker, Weijian Shan and Dong-Jae Kim (1994). "Platform technologies and national industrial networks" (Philadelphia: Wharton School, University of Pennsylvania), mimeo.

Krugman, Paul R. (1992). "A global economy is not the wave of the future", *Financial Executive* (March/April), pp. 10-13.

Lawrence, Robert Z. (1995). "Emerging regional arrangements: building blocks or stumbling blocks?", in Jeffrey A. Frieden and David A. Lake, eds., *International Political Economy* (New York: St. Martin's Press), pp. 407-415.

Mytelka, Lynn (1991). "Crisis, technological change and the strategic alliance", in Lynn Mytelka, ed., *Strategic Partnerships and the World Economy* (Rutherford, N.J.: Fairleigh Dickinson University Press), pp. 7-34.

Negroponte, Nicholas (1995). *Being Digital* (New York: Knopf).

Orwell, George (1949). *1984* (New York: New American Library-Signet Classic).

Ruggie, John Gerard (1993). "Territoriality and beyond: problematizing modernity in international relations", *International Organization*, 47, 1 (Winter), pp. 139-174.

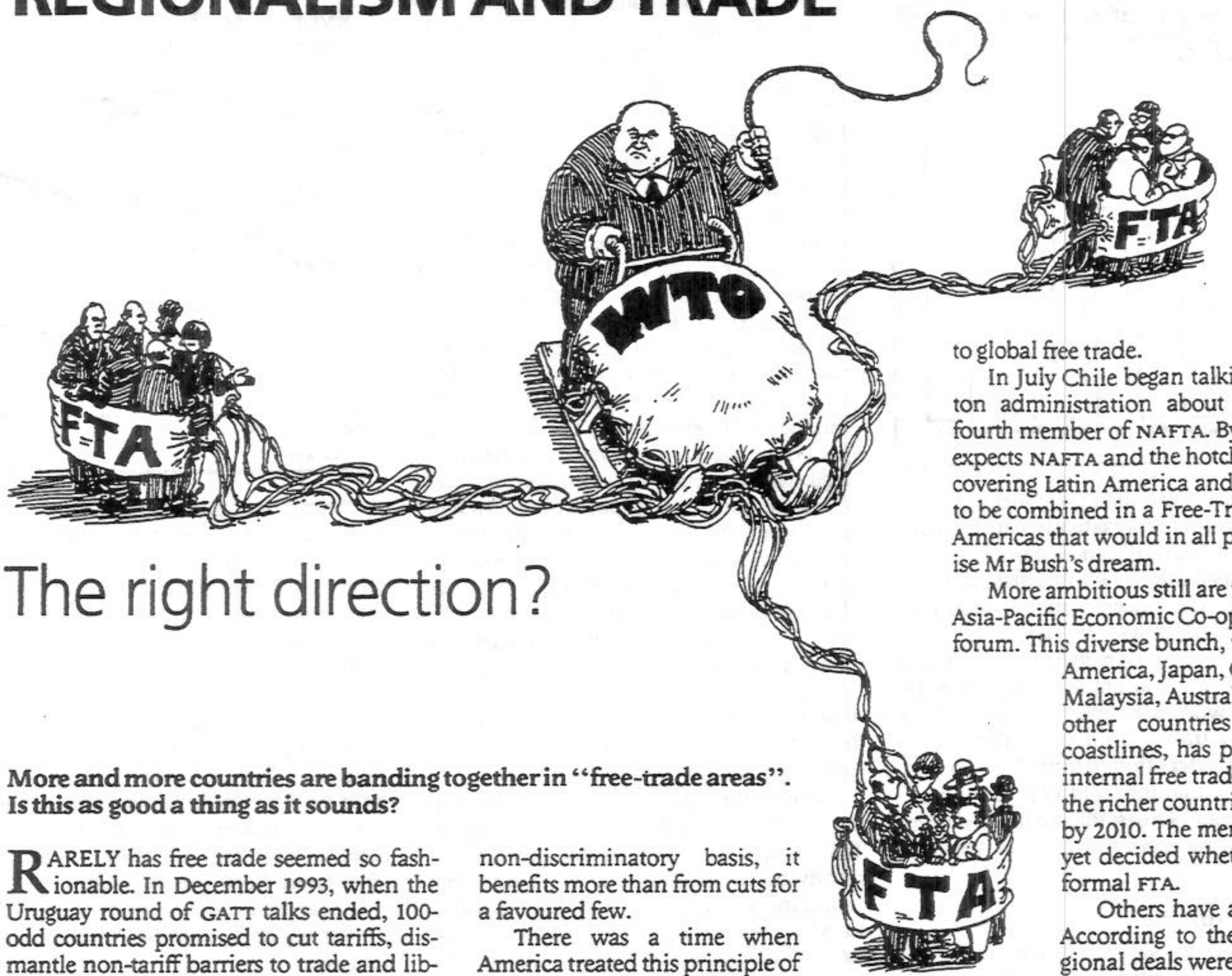
United Nations Conference on Trade and Development, Division on Transnational Corporations and Investment (UNCTAD-DTCI) (1993). *World Investment Report 1993: Transnational Corporations and Integrated International Production*. Sales No. E.93.II.A.14.

———— (1994). *World Investment Report 1994: Transnational Corporations, Employment and the Workplace.* Sales No. E.94.II.A.14.

United States, Office of Technology Assessment (1993). *Multinationals and the National Interest: Playing by Different Rules* (Washington, D.C.: U.S. Government Printing Office).

Witherell, William H. (1995). "The OECD Multilateral Agreement on Investment", *Transnational Corporations*, 4, 2 (August), pp. 1-14.

REGIONALISM AND TRADE



The right direction?

More and more countries are banding together in "free-trade areas". Is this as good a thing as it sounds?

RARELY has free trade seemed so fashionable. In December 1993, when the Uruguay round of GATT talks ended, 100-odd countries promised to cut tariffs, dismantle non-tariff barriers to trade and liberalise trade in services. As if that were not enough to be going on with, scarcely a week passes without some politician, somewhere, mooting a new effort to make trade, as he sees it, freer still.

A lot more often than not, such efforts centre on agreements to liberalise trade on a regional basis. Free-trade areas (FTAs, in which partners agree to scrap tariffs on trade among themselves) and customs unions (FTAs whose members have common tariff rates on imports from non-members) have been sprouting all over the globe.

Unsurprisingly, many think that this mixture of modish enthusiasms is all to the good. But dedicated followers of fashion are divided. While some claim to see regionalism and globalism as complementary, others see them as contradictory. These critics say that, belying their appealing name, FTAs will make world trade less free.

The basic principle behind the World Trade Organisation (WTO), the newly minted guardian of the good that GATT has wrought, is that member countries should not discriminate between one trading partner and another. If a country cuts tariffs on trade with one country, it must cut tariffs for all. This principle is well grounded in economic theory. If a country cuts its tariffs on a non-discriminatory basis, it benefits more than from cuts for a favoured few.

There was a time when America treated this principle of non-discrimination as an article of faith. That faith was a robust one. Although the theologians of GATT would allow FTAs and customs unions under various conditions, America would have none of it. The first signs of doubt came in 1982. William Brock, then America's chief trade negotiator, became frustrated by the refusal of other countries to agree to a new round of GATT talks. He decided that America might make regional deals after all.

Everybody's doing it

Now the United States has taken up regionalism with a convert's zeal. An FTA between America and Canada was created in 1988; the North American Free-Trade Agreement (NAFTA), a three-way pact including Mexico as well as Canada, followed in 1993. America got most of what it wanted in the 1980s with the completion of the Uruguay round; even so, it shows no sign of returning to the globalist fold. The enthusiasm George Bush voiced for an FTA stretching "from Alaska to Tierra del Fuego" persists. Jagdish Bhagwati, an economics professor at New York's Columbia University, derides the Clinton administration for its "infatuation" with FTAs. If he and his friends are right, this enthusiasm presents a real danger to global free trade.

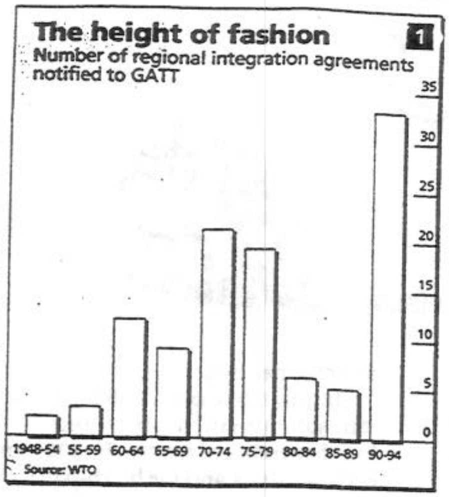
In July Chile began talking to the Clinton administration about becoming the fourth member of NAFTA. By 2005 America expects NAFTA and the hotch-potch of FTAs covering Latin America and the Caribbean to be combined in a Free-Trade Area of the Americas that would in all probability realise Mr Bush's dream.

More ambitious still are the plans of the Asia-Pacific Economic Co-operation (APEC) forum. This diverse bunch, which includes America, Japan, China, Taiwan, Malaysia, Australia and a dozen other countries with Pacific coastlines, has pledged itself to internal free trade by 2020, with the richer countries getting there by 2010. The members have not yet decided whether to create a formal FTA.

Others have also been busy. According to the WTO, 109 regional deals were notified to the GATT between 1948 and the end of last year. Nearly one-third of those were signed between 1990 and 1994 (see chart 1 on next page). January 1st this year saw the accession to the European Union (EU), the world's biggest trading block, of Austria, Finland and Sweden, and the creation of Mercosur, a customs union between Argentina, Brazil, Paraguay and Uruguay.

More agreements are on the way. A South Asian Preferential Trade Area has been promised. The seven members of the Association of South-East Asian Nations have agreed to set up an FTA by 2003. On September 11th it was agreed that Slovenia would join the Czech Republic, Hungary, Poland and Slovakia in the Central European Free-Trade Area (see page 60). Next January 1st, the EU and Turkey are due to set up a customs union. The EU is expected to add to its free-trade deals with North African countries; it already has accords, called the Europe Agreements, with several formerly Communist states in Central and Eastern Europe.

The grandest idea of all is for an FTA between North America and the EU. The loudest cooings have come from Klaus Kinkel, Germany's foreign minister, and from Canada's trade minister, Roy MacLaren. For


The height of fashion 1

Number of regional integration agreements notified to GATT

Source: WTO

now, America and the EU are sticking to less formal ideas—a "transatlantic business dialogue", or a "transatlantic economic space". But one day a full-blown FTA may follow.

To see these developments purely in terms of trade is wrong. The EU is much more than a trade agreement. APEC has as much to do with maintaining security in the Pacific as with economic initiatives. And moves for a transatlantic trade alliance are part of a broader effort to bind America and Europe now that the Cold War has passed. Yet their creators insist that all this togetherness will bring economic benefits, too. Are they right?

Hold on a minute

Economic theory says that the effects of regional deals are ambiguous. The traditional analysis is based on a paper written in 1950 by Jacob Viner, a Canadian economist. Suppose that two countries agree to scrap their tariffs on each other's goods, but keep them on imports from third countries. They make their citizens better off by "creating" trade: both countries import from their neighbours goods that they used to make less cheaply themselves. But at the same time some trade is "diverted": the two countries now buy from one another things that they used to import from more efficient producers in other countries. That makes them (as well as those producers) worse off in a variety of ways, most notably through lost tariffs and lost opportunities.

Paul Krugman, a Stanford University economist, has suggested that regional FTAS may do more creation than diversion because they are "natural". His argument is that, although trade is based on comparative advantage, it is influenced by the realities of geography. Since near-neighbours usually trade a lot with each other anyway, an FTA between them will divert little. In contrast, FTAS between countries that are far apart—on different continents, say—are "unnatural": if they form an alliance, they will probably succeed in creating little trade while diverting plenty.

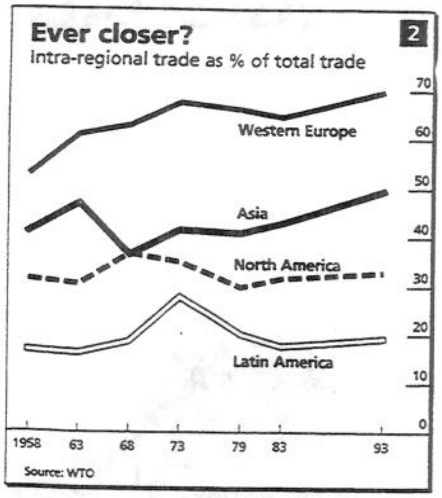
A report by economists at the WTO seems to concur, in that it argues that regional agreements have so far done little damage*. The creation of the European Union (EU), the world's biggest and most cohesive trade block, has led to a big increase in intra-regional trade in Western Europe (see chart 2). That may reflect some diversion, but the WTO points out that barriers to trade with outsiders have also come down. Trade with non-EU countries has risen as a proportion of GDP, even though it has fallen as a proportion of total trade.

Jeffrey Frankel of the Institute for International Economics (IIE), a Washington, DC, think-tank, Ernesto Stein of the Inter-American Development Bank and Shang-Jin Wei of Harvard University think intra-regional trade may not be so nicely natural: it may be sinisterly "supernatural". In a recent paper† the three looked at trade among 63 countries between 1965 and 1990, comparing real trade with predictions based on the size of the economy, geographical distance, common borders and languages and so on. They found that trade within the EU, Mercosur and the Andean Pact is larger than they would expect—thus supernatural. In 1990 Mercosur's intra-regional trade was eight times bigger than expected.

This crime against nature might be seen as evidence that trade diversion is going on. But though that may be so, the story is clearly more complicated. The researchers found intra-regional biases in East Asia and in APEC, where there are no FTAS as yet. Arvind Panagariya, an economist at the University of Maryland, argues that extra-regional biases also abound. Most countries, he says, trade more than geography alone would predict with North America, East Asia and some EU countries.

As well as opening the possibility of trade diversion, FTAS can also do less than would be hoped for in trade creation. Take "rules of origin", which are used in FTAS to determine when goods qualify for tariff-free treatment. If a car made in Canada with some Japanese parts counts as Canadian, it can be exported to America tariff-free; if it does not, it cannot. As Mr Bhagwati points out, these rules are a nonsense in a globalised economy, in which goods are made from bits made in many countries: it is daft to try and sort out which goods are whose. As is often the way with nonsense, the rules are also fiendishly complicated: NAFTA's run to 200 pages.

In 1992 America and Canada fell out over the American content of Canadian-built Hondas. The Canadians, reasonably, argued that all their American-made parts counted; but the Americans insisted that only the American-made element of the American parts did. Not much trade created there; and rules of origin clearly divert trade, too. They can encourage Canadian firms to switch from Japanese parts to American ones.


Ever closer? 2

Intra-regional trade as % of total trade

Source: WTO

Mr Panagariya adds another twist to Viner's tale. If a country with fairly high tariffs (Mexico, say) enters an FTA with a more open country which supplies a lot of its imports (America, for instance) it loses tariff revenues. Thanks to the tariff on imports from third countries elsewhere, it is possible that some import prices will not fall. In that case, the tariff income would, in effect, be transferred from the Mexican treasury to American firms. Mexican firms will get no similar windfall, because American tariffs are already low.

Best foot forward . . .

These problems should give the designers of regional FTAS pause. Though trade diversion may not be measurable, it is hurting people inside and outside the FTAS. But the biggest objection to their designs is not a matter of technicalities and means, but of principles and ends. As their critics see it, people putting together regional FTAS are, as a result, less likely to fight for global free trade. That is the real problem.

The supporters of FTAS say that they share the goal of freer world trade. Fred Bergsten, head of the IIE, argues that FTAS will help bring about global free trade sooner, by keeping up the momentum at a time when, all passion spent on the Uruguay round, global negotations are languishing. With the EU taking time off to grapple with next year's intergovernmental conference, monetary union and other pressing matters, he sees regional agreements as the best outlet for America's free-trade aspirations. For Mr Bergsten, that means APEC: he is chairman of the Eminent Persons Group, a committee advising the 18 governments.

There are three ways in which FTAS might help, say proponents such as Mr Bergsten. First, regional deals can kick off the liberalisation of things not yet covered, or covered only poorly, by the WTO. The protection of foreign direct investment is probably the most important of these at the moment: there are WTO rules, but they are weak.

---

*"Regionalism and the World Trading System". April 1995

†"Trading Blocs and the Americas: The Natural, the Unnatural and the Supernatural". Forthcoming in the Journal of Development Economics

A second argument is that regional forums, because they involve fewer countries, are more flexible. "You can turn on a dime," as one seasoned American official puts it. Problems that would take years to sort out in global talks, he says, can be dealt with in months in smaller groups.

And third, some argue that regional trade deals "lock in" liberalisation, especially in developing countries. On this argument, by joining NAFTA Mexico made a commitment, enshrined in a treaty, that it would not unwind its recent economic liberalisation.

These arguments may hold—but only up to a point. To be sure, NAFTA's rules go further than the WTO's in at least two respects. Its investment-protection code allows firms, as well as governments, to bring cases against host governments before dispute-settlement tribunals. And services trade between NAFTA countries is liberalised except in specifically excluded sectors; in the WTO, only trade in named areas will be freed. Both are models that the WTO would do well to copy.

Not all FTAs and proto-FTAs are as well equipped, though. The investment code for APEC, unlike that for NAFTA, is flimsy and non-binding. And the fact that there are arguments in NAFTA's favour should blind no one to the arguments against. As part of its admission fee, Mexico had to sign side agreements to maintain labour and environmental standards. The Clinton administration is adamant that Chile must do the same; naturally, the Mexicans agree. Some Democrats would like to add obligations to curb drug trafficking or conditions on the treatment of indigenous peoples.

This ability to boss people around is one of the political enticements of FTAs; it is part of the economic case against them. Forcing poorer countries to improve their labour and environmental standards as a precondition for trade risks making them poorer still. It often raises costs in the industries in which they have a comparative advantage. And some who demand such standards are really seeking protection, not a cleaner planet or better treatment for foreign workers. That suits America's labour lobby and some businesses, but does little for trade.

Nor is it clear that regional agreements necessarily bring about reform more quickly than global talks. A lot might be achieved globally before APEC's notional free-trade date of 2020. Although the Uruguay round took seven years to complete (and bits of it are still unfinished) it did end up reforming 100 countries. The FTA between America and Canada, with just two participants, took three years to negotiate; Mexico applied to join the club in June 1990, some three-and-a-half years before the NAFTA agreement finally squeaked through Congress. Now congressional Republicans say that they will not give Mr Clinton "fast-track" negotiating authority, under which lawmakers would forgo the right to amend legislation implementing Chile's accession, as long as he insists on the extra clauses. This stand-off could stall the expansion of NAFTA.

The third argument in favour of regional deals, that they can help to lock in reforms, is spurious. Did NAFTA really do that for Mexico? The reverse seems more plausible: that the Americans saw NAFTA as worth negotiating because Mexico had already liberalised much of its economy unilaterally. The same surely applies to Chile. And it is hard to see what makes regional agreements special: WTO commitments are equally credible, without discriminating between trading partners.

Tests of faith

If FTAs are really going to help world trade they must try to be similarly non-discriminatory. Entry to FTAs should be granted willingly, not grudgingly. Members of FTAs should commit themselves to reducing their external, as well as internal, barriers to trade. Like the Marxist state, they should be committed to withering away.

Judged by these standards, FTAs are doing poorly. The EU has been in no hurry to admit East European countries as members. America treats entry into NAFTA as a reward for carrying out the right economic policies rather than a mutually helpful agreement. The jury on APEC is still out. There is a school of thought within the organisation that favours freeing trade with non-members as well as with members. If it prevails, free traders can afford a smile. Others hold that APEC nations should liberalise among themselves, and use the promise of opening their markets as a way of prising open markets elsewhere—ie, in the EU.

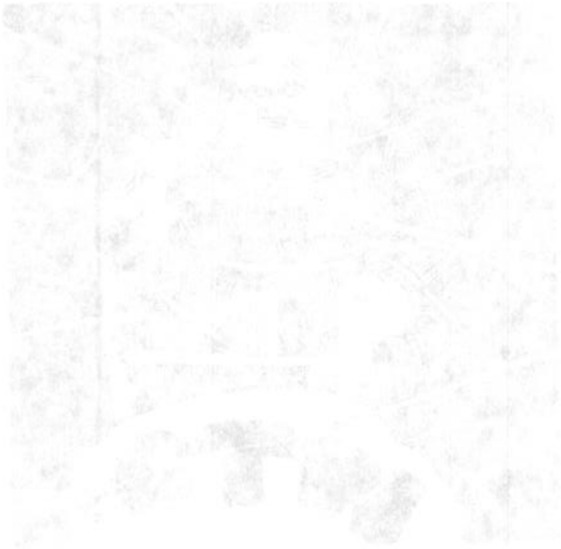
That would be bad. Thinking of free trade as a concession to others, rather than a boon for one's own citizens, is a cardinal error in trade diplomacy. Both Europe and APEC countries would benefit from freer trade between the two regions. True, it would be better *if* both liberalised straight away. But both would also gain if either lifted their trade barriers unilaterally. While the bargaining went on, trade between Europe and the Pacific would be held back, with both losing out.

The WTO ought to be trying to make FTAs as non-discriminatory as possible. But its rules do not give it clear enough authority. Article XXIV of the GATT permits FTAs and customs unions so long as they cover "substantially all trade" and do not increase external trade barriers "on the whole". This has proved as impossibly vague as it sounds. Of 69 GATT working parties set up to examine the consistency of FTAs and customs unions with Article XXIV, only six managed to reach a conclusion.

So how can the WTO become a better overseer? Its recent report suggests three possibilities. First, Article XXIV could be tightened up. For instance, countries forming an FTA could be required to cut their tariffs to the lowest level of any member. Second, FTA members should be required to notify the WTO before signing their agreements, so that working parties could deliberate before the agreements came into force. Third, the monitoring of FTAs could be improved.

All these are worth pursuing. The first would guarantee some external liberalisation: some EU tariffs would have been cut when new members joined at the start of this year, Mexico would eventually have to lower its tariffs to American levels. The second would probably reduce foreseeable trade diversion and encourage the use of liberal accession clauses. The third might keep FTAs in line.

But none of these safeguards will be set up unless the WTO's member states (that is to say, the very governments who create FTAs) truly want to commit themselves to freeing trade multilaterally. So far, many aspects of recent FTAs—the rules of origin, NAFTA's labour and environment clauses—suggest that less wholesome motives are shaping their policies. If freer trade is to become more than just fashionable talk, that will have to change. ■



45

LIPPINCOTT LIBRARY
DOCUMENT DELIVERY

WHARTON REPROGRAPHICS

The Free Trade Area of the Americas: Good for Latin America?

Arvind Panagariya

1. INTRODUCTION

EXCLUDING Mexico, Latin America accounts for less than 2.5 per cent of the world trade. Is the creation of a trading bloc between this economically small region and the members of the North American Free Trade Agreement (NAFTA) a good idea? Many analysts answer, yes, to this question arguing that even though the effect of the bloc on the United States and Canada is likely to be small, it will bring large gains to Latin America. Afterall, for Latin America, trading freely with the United States is like trading freely with the world. And if free trade is supposed to bring gains, the Free Trade Area of the Americas (FTAA) must surely do so.

Unfortunately, this argument is based on flawed reasoning. In the policy debate, there is a strong tendency to equate free trade areas (FTAs) with free trade. But as Viner (1950) taught us many decades ago, this equation is erroneous: FTAs liberalise trade on a discriminatory basis by removing trade barriers exclusively against union members while free trade requires a removal of the barriers against all trading partners. Indeed, discriminatory liberalisation under FTAs may *increase rather than decrease* protection and may, indeed, make the countries undertaking the liberalisation as well as the world as a whole worse off. The conventional wisdom which emerged from Viner's seminal contribution was that, on balance, FTAs were likely to be harmful; the United States resisted all calls for the formation of trading blocs in the post-Second-World-War period.[1]

ARVIND PANAGARIYA is Professor of Economics and Co-director, Center for International Economics, University of Maryland, College Park. He is indebted to two referees, Jagdish Bhagwati and Dani Rodrik for comments on this paper and to John Nash, Maurice Schiff, Sethaput Suthiwart-Narueput and Wendy Takacs for suggestions on related work which helped to improve this paper. Thanks are also due to Gabriel Castillo for valuable research assistance.

[1] The United States supported the formation of the European Economic Community (EEC) purely for political reasons. At that time, the United States saw a united Western Europe essential to meet the growing Soviet threat. See Bhagwati (1993) in this context.

© Blackwell Publishers Ltd 1996, 108 Cowley Road, Oxford OX4 1JF, UK
and 238 Main Street, Cambridge, MA 02142, USA.

The World Economy, Volume 19, Number 5, Sept. 1996

In the mid-1980s, for a variety of reasons discussed in Bhagwati (1993 and 1995), the United States turned to regionalism.[2] The Canada-USA Free Trade Agreement (CUFTA) was concluded in 1988 and NAFTA in 1992. Most recently, in 1994, an agreement was signed to create by the year 2005 a Free Trade Area of the Americas (FTAA) consisting of the entire Western Hemisphere except Cuba. Alongside these developments, the conventional wisdom on FTAs has been shifting. Such distinguished economists as Dornbusch (1993), Krugman (1991 and 1993) and Summers (1991) have enthusiastically embraced regionalism as a principal instrument of free trade.[3] Summers (1991), a staunch supporter of trading blocs, has taken the following position:

> (E)conomists should maintain a strong, but rebuttable, presumption in favor of all lateral reductions in trade barriers, whether they be multi, uni, bi, tri, plurilateral. Global liberalization may be best, but regional liberalization is very likely to be good.

In this paper, I take issue with the advocates of NAFTA and FTAA and argue that the key issue is not which lateral is chosen but whether the given lateral is implemented on a discriminatory or nondiscriminatory basis. My central thesis is that when trade liberalisation is discriminatory as in an FTA, within the Vinerian framework, the effects of freer trade have a strong mercantilist bias: a country benefits from *receiving* a preferential (or discriminatory) access to the partner's market and is hurt by *giving* the partner a similar access to its own market. When the country gives access to the partner on a preferential basis, it loses the tariff revenue collected on imports from the partner. The revenue goes to boost the terms of trade of the latter. The reverse happens when the country receives a preferential access from the partner. On balance, then, the country which liberalises most, is likely to lose. In the context of NAFTA and FTAA, the United States and Canada already have virtual free trade on a nondiscriminatory basis while Latin America has high tariffs. Therefore, within NAFTA and FTAA, Mexico and Latin America will give a much greater discriminatory access to the United States and Canada than they will receive from the latter. The inevitable conclusion is that the static welfare effects of NAFTA and FTAA on Mexico and Latin America will be negative.

Of course, the traditional Vinerian analysis is insufficient to establish the case against NAFTA and FTAA from the viewpoint of Mexico and Latin America. Recently, it has been argued that the gains to a developing country from an FTA with a large, rich trading partner go far beyond the traditional static effects. These

---

[2] The principal reason why the United States, an ardent supporter of the nondiscriminatory process, turned to preferential trading arrangements was the failure to secure an agreement from Europe and developing countries to start a new round of multilateral trade negotiations.

[3] A referee of this journal has pointed out that 'from recent statements by Messrs. Krugman and other regional enthusiasts at least the emphasis is gone and newly found enthusiasm for the WTO can be detected.' In the case of Krugman, this observation is supported by a recent presentation by him at the 1996 American Economic Association meetings in San Francisco.

© Blackwell Publishers Ltd 1996

new, *non-traditional* gains arise from a guaranteed access to the large market of the rich partner, protection from predatory actions such as anti-dumping by the latter, a 'lock' on the country's own reforms, and a steady flow of foreign capital. If these gains are sufficiently large, FTAA may be desirable despite adverse welfare effects in static terms.

In the following, I will examine these non-traditional gains critically and argue that, in the absence of any means of verification, their significance was greatly exaggerated by the proponents of NAFTA. The latter often claimed gains from NAFTA that simply did not exist and disregarded benefits of the General Agreement of Tariffs and Trade (GATT) — now the World trade Organisation (WTO) — that did exist. In this context, it is critical to distinguish between the rules of the game in the European Union and those in NAFTA.[4] The European Union (EU) and NAFTA are fundamentally different institutions: the former is a customs union with high degree of capital mobility and a commitment to eventual labour mobility. The latter is an FTA with complex rules of origin and essentially hostile to labour mobility.

The remainder of the paper is organised as follows. In Section 2, I outline the basic Vinerian economics using a simple partial equilibrium model. In addition to pointing out the relevance of the mercantilist approach for the economics of FTAs, I also question the analytic basis of some key justifications for FTAs in recent writings. My analysis is not new; it derives from the basic Vinerian model and can be found in different forms in the literature. The paper's contribution lies in reminding readers of the forgotten lessons in the context of NAFTA and FTAA.

In Section 3, I discuss some aspects of regional arrangements between developed and developing countries — the so-called North-South arrangements — not covered by the Vinerian analysis. In Section 4, I subject NAFTA to a critical scrutiny and, in Section 5, I compare it to the EU. In Section 6, I take a close look at other Latin American countries and argue that, on balance, they will lose from FTAA. In Section 7, I address issues related to policy harmonisation. My conclusion here is that policy harmonisation within North-South regional schemes is a costly affair for southern countries. For one thing, as has been systematically shown in various contributions in Bhagwati and Hudec (1996), there is no compelling case for policy harmonisation, specially when countries are disparate in terms of size, income levels and preferences. Moreover, even if a plausible case could be made for harmonisation, it requires a global context.[5] Harmonisation by southern partners in North-South schemes towards northern

---

[4] By core members in the EU, I mean full members of whom there are 15 currently. Core members are to be distinguished from associate members, mainly countries in Central and Eastern Europe, for whom the rules of the game are different.

[5] Bhagwati (1995) has argued persuasively that as far as labour and environmental standards are concerned, a case for harmonisation can simply not be made.

© Blackwell Publishers Ltd 1996

countries' standards will leave the former at a disadvantage *via-a-vis* their other southern competitors in regions such as East Asia. A summary of the main conclusions is provided in Section 8.

2. MERCANTILIST BIAS IN THE EFFECTS OF DISCRIMINATORY LIBERALISATION[6]

In a small open economy, and up to a point, even in a large open economy, the primary source of gains from trade liberalisation is the reduction in home tariffs. Reduced protection at home leads to a more efficient allocation of resources and of consumer expenditures. Yet, in trade negotiations, the reduction in home tariffs is viewed by negotiators as a cost and a reduction in the partner country's tariffs as benefit. Because of this mercantilist bias, it is commonly suggested that the negotiations under the General Agreement on Tariffs and Trade (GATT) lead to the right outcome for the wrong reason.

Surprisingly, when it comes to regional integration, at least within the Vinerian framework, the mercantilist approach is right: gains from regional integration come primarily from a reduction in the partner country's tariffs. Though this point is implicit in the standard Vinerian analysis, to my knowledge, it has not been recognised explicitly. I will first explain the point within a simple model and then deal with complications which may undermine it.

a. Preferential Tariff Liberalisation: The Basic Case

Consider a world consisting of three countries: A B and C. Think of A as Mexico which has high tariffs initially. In Figure 1, M_AM_A represents the import demand for a product by A. The height of this curve represents the marginal benefit from imports and, hence, the area under it the total (gross) gains from trade.

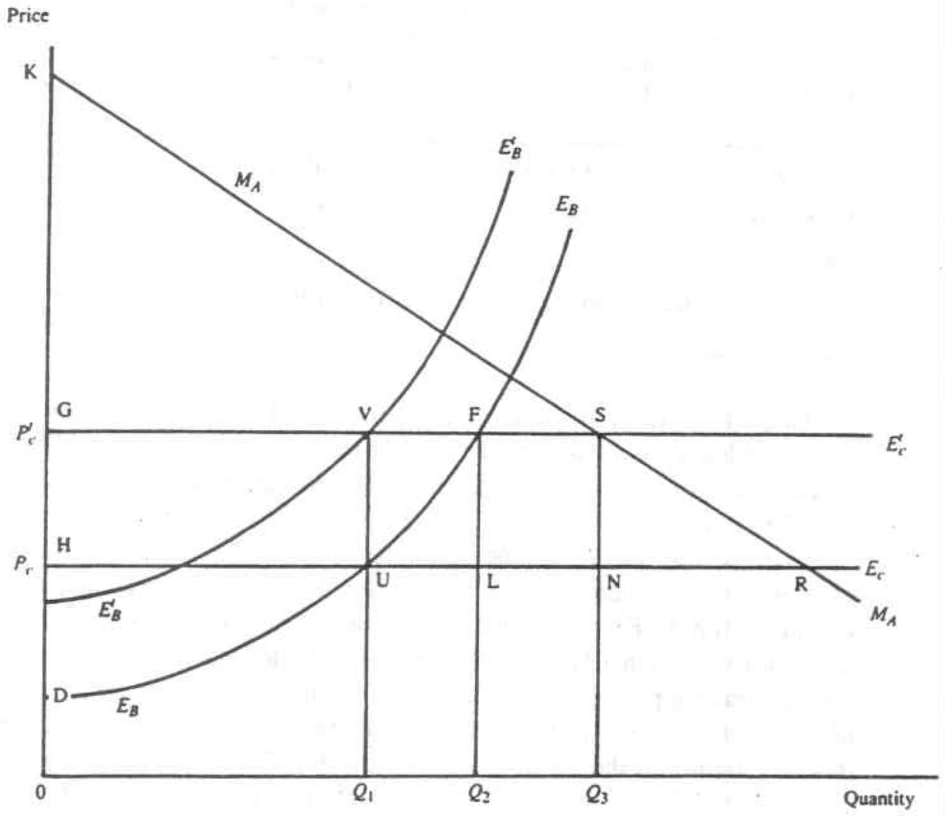
Before we introduce export supplies of B and C, recall that Viner introduced the concepts of trade creation and trade diversion using a case in which supply prices of both B and C are constant. An important drawback of this case is that, in equilibrium imports come from only one country. That is, whichever country has a lower tariff-inclusive price supplies *all* imports. A more realistic case is the one in which imports come from both countries. This case is captured most simply under the assumption that the supply price rises with exports for one country, say B, and remains constant for the other, C. As I will discuss later, this case, too, has certain limitations but it captures the essence of our argument without requiring us to deal with the complications of the large-union case.

---

[6] This section and Section 6 draw on Panagariya (1995).

© Blackwell Publishers Ltd 1996

FIGURE 1



In Figure 1, I represent the export supply of B by the upward sloped curve E_BE_B and that of C by the horizontal line P_CP_C. Under a nondiscriminatory tariff at rate t per unit, supplies from B and C, as perceived by buyers in A, are given by $E^t_BE^t_B$ and $P^t_CP^t_C$, respectively. Total imports into A equal $0Q_3$ of which $0Q_1$ come from B and Q_1Q_3 from C. A collects tariff revenue in the amount represented by rectangle GHNS. The gains from trade amount to the area under the import-demand curve and above the domestic price plus the tariff revenue, i.e., triangle KSG plus rectangle GHNS. For country B, the gains from trade equal the area above E_BE_B and below the net price received, P_C, and equals HUD. Country C neither gains nor loses from trade. Table 1 summarises this information in column 1.

Suppose now that A decides unilaterally to adopt a policy of nondiscriminatory free trade. The price in A declines to P_C, imports from B do not change, and imports from C rise by NR. Tariff revenue disappears but the gains from trade rise to KGS+GHNS+RSN. There is a net welfare gain of RSN. The extra gain comes from increased benefits to consumers. The gains to country B remain

© Blackwell Publishers Ltd 1996

TABLE 1
Gains From Trade Under Unilateral Liberalisation and FTA
(The Table Relates to Figure 1)

Policy *Country*	*Nondiscriminatory Tariff at Rate t (1)*	*Complete Free Trade (2)*	*FTA with B (3)*
Country A	KGS + GHNS	KGS + GHNS + RSN	KGS + GHNS − GFLH
Country B	HDU	HDU	HDU + GFLH − UFL
Country C	0	0	0
World	KGS + GHNS + HDU	KGS + GHNS + HDU + RSN	KGS + GHNS + HDU − UFL

unchanged at HDU. Because of the perfectly elastic supply, country C neither gains nor loses from trade before or after trade liberalisation by A. Therefore, the world as a whole benefits by area RSN. These changes are summarised in column 2 of Table 1.

Next, assume that A has the option to form an FTA with B or C. It should be obvious that the outcome of the FTA will depend critically on whether the partner country is B or C. For the moment, I assume that the partner country is B and return to the other case later. Because the rest of the world is likely to be much larger than a single trading partner, *prima facie*, it is not unreasonable to assume that the rest of the world's supply is more elastic than that of the partner. Figure 1 represents this assumption in the extreme form by making C's supply perfectly elastic.

Under the FTA arrangement, A eliminates entirely the tariff on B but retains it on C. Imports from B rise to $0Q_2$ and those from C decline to Q_2Q_3. Though there is trade diversion, B gains from the FTA due to an improvement in its terms of trade. The net price received by the exporters of B increases from P_C to P_C^t and the gains from trade to B rise to HDU+HUFG. *Country B gains from A's liberalisation precisely as mercantilists would predict.*

Because imports continue to come from C before as well as after the FTA and C's supply is perfectly elastic, the price in A is unchanged. But now that there is no tariff revenue on goods coming from B, A's gains from trade decline by GFLH. *Once again, as mercantilists would have argued, A loses from its own liberalisation*. Because the FTA diverts imports Q_1Q_2 from the more efficient C to less efficient B, A's loss exceeds B's gain by area UFL. The world as a whole loses by area UFL. The last column in Table 1 shows these changes.

In his influential paper mentioned earlier, Summers (1991) chastises economists for paying too much attention to second best considerations while evaluating regional arrangements. Referring to trade diversion, he writes

> I find it surprising that this issue is taken so seriously — in most other situations, economists laugh off second best considerations and focus on direct impacts.

© Blackwell Publishers Ltd 1996

Taking issue with Summers' position in this and the earlier quotation in the Introduction, Bhagwati (1995) has recently retorted thus,

> My gut reaction is to wonder what Summers would say if, on hearing that he wanted our spending to be cut, I told him, 'Cut anything you want, wherever you wish, it does not matter what, as it all leads to reduced spending.' Or, on hearing that he wanted revenues to be raised, I advised him, a wonderful public finance specialist of the neoclassical variety, to raise any and every tax in any way that he could since all taxes raised revenue.

The analysis in this section complements Bhagwati's argument. The loss to A from its own liberalisation arises primarily from the 'impact effect' of the FTA rather than second-best considerations. If we assume that the initial imports from the trading partner are large, the loss due to trade diversion — triangle UFL in Figure 1 — accounts only for a small fraction of A's loss. The large part of the loss — rectangle GVUH — is accounted for by the lost tariff revenue on the original imports from the partner. In denouncing trade diversion as a major concern, Summers has clearly overlooked the fact that FTAs can also give rise to *large redistributive effects* between countries and that, unlike efficiency effects which are triangles, redistributive effects are rectangles!

My analysis also raises doubts about the argument injected into the debate recently that countries which trade with each other a lot are 'natural' trading partners and regional arrangements among them are beneficial. It is not clear what it means to be 'natural' trading partners but a quotation from Summers (1991) may help:[7]

> Are trading blocs likely to divert large amounts of trade? In answering this question, the issue of natural trading blocs is crucial because to the extent that blocs are created between countries that already trade disproportionately, the risk of large amounts of trade diversion is reduced.

Drawing upon the analytic results from the past literature, particularly Lipsey (1958), Bhagwati (1993) has questioned the presumption that trade diversion is minimal if potential members trade disproportionately more with each other. My analysis adds to Bhagwati's critique of the 'natural trading partners' argument. It is evident from Figure 1 that the larger the initial quantity of imports from a trading partner, the greater the loss to the country liberalising preferentially from a given trade diversion. That is to say, the more natural the trading partner

---

[7] As far as I am able to trace, the idea originated in Wonnacott and Lutz (1989). Krugman (1991) also makes this argument as follows, 'To reemphasize why this matters: if a disproportionate share of world trade would take place within trading blocs even in the absence of any preferential trading arrangement, then the gains within trading blocs even in the absence of any preferential trading arrangement, then the gains from trade creation within blocs are likely to outweigh any possible losses from external trade diversion.' Recently, Primo Braga, Safadi and Yeats (1994) have pushed the 'natural trading partners' argument one step further by suggesting that, having become more interdependent in recent years, Latin American economies are more ripe for regional integration among themselves today than in the 1960s and 1970s. As discussed in the text, economic theory offers no basis for this conclusion.

© Blackwell Publishers Ltd 1996

according to Summers' definition, the larger the loss from a discriminatory trade liberalisation with it!

Finally, it is being suggested lately that given lower levels of trade restrictions than in the 1960s and 1970s, the chances that trade diversion will dominate trade creation are low. This is exactly the opposite of the conclusion coming out of the Vinerian analysis. Thus, in Figure 1, if the initial nondiscriminatory tariff is sufficiently high, an FTA between A and B eliminates C as a supplier of the product. In this case, the FTA lowers the internal price in A and gives rise to trade creation. Under some conditions, this trade creation can outweigh the tariff-revenue loss and may improve welfare. By contrast, if the initial tariff is low, the chances that the formation of the FTA will eliminate imports from C and lower the internal price are poor.

b. Beyond the Basic Case

The conclusion that A's preferential liberalisation hurts itself and benefits its union partner has been derived under the assumption that the supply of B is less than perfectly elastic and that of C is perfectly elastic. In this setting, the union partner is a less efficient supplier of the product than the outside world. What will happen if the situation was reversed and A formed a union with C rather than B?

In Figure 1, a union between A and C lowers the price in A to P_C. Though there is no gain to the union partner, A's gain from the FTA exceeds that under nondiscriminatory liberalisation. Relative to the initial, nondiscriminatory tariff, A gains area SNR plus the tariff revenue (not shown) on imports from the outside country B. We are back to the neoclassical world where A's liberalisation benefits itself and not B.

This case clearly undercuts the arguments made in Section 2*a*. Therefore, it is important to ask how relevant it is empirically. It is perhaps reasonable to assert that a union partner is likely to resemble B for some products and C in other products and, therefore, the effect of the FTA will be ambiguous in general. In the specific case of NAFTA, a common claim has been that, on net, Mexico is likely to gain because the United States is very large and, *therefore*, the most efficient supplier of a majority of Mexico's products.

There are at least two reasons why this conclusion is unwarranted. First, given that the outside world includes the EU, Japan, China, Korea, Hong Kong and numerous other highly competitive countries, the conclusion that the United States and Canada are the most efficient suppliers of a large majority of Mexico's products is highly suspect. Indeed, if it were true, we will be hard-pressed to explain the persistent demands for anti-dumping and other forms of protection in the United States. Second, recall that if the union partner is a large supplier of imports, the losses to A in the case of trade diversion are large and the gains to it in the case of trade creation are small. Therefore, even if the union partner is the

© Blackwell Publishers Ltd 1996

most efficient supplier of the majority of A's imports, losses may outweigh the gains. In the case of NAFTA, the United States does account for a sufficiently large proportion of Mexico's imports that the losses to the latter in trade-diversion cases will outweigh the gains to her in trade-creation cases.

Before I conclude the discussion of static welfare effects of an FTA, two additional possibilities deserve to be considered: (i) export-supply curves are upward sloped for both B and C and (ii) exports of B and C are imperfect substitutes. In both of these cases, the small-union assumption is violated and a complete elimination of the tariff by A, whether on a discriminatory or nondiscriminatory basis, is not the optimal policy.

Bearing this fact in mind, let me note that the analysis of Section 2*a* remains valid under the following circumstances. In case (i), if the elasticity of supply of the outside country is high in relation to that of the union partner, a discriminatory tariff reduction is likely to hurt even when a nondiscriminatory reduction will not. The discriminatory tariff reduction will benefit B. In case (ii), if the imports coming from the union partner and the outside country are sufficiently close substitutes, discriminatory liberalisation by A will hurt that country and benefit the union partner. Because neither of these cases are implausible in reality, I conclude that the presumption is in favour of the conclusion, derived earlier, that within a discriminatory liberalisation scheme, a country is hurt by its own liberalisation and benefits from the partner's liberalisation.[8]

c. FTAs Under Quotas

Let us now briefly turn to the case of a quota. The first point to note is that under a quota, the wedge between the domestic and the border price — the so-called quota rent — is captured by the recipient of the licence. For example, if import licences are issued free of charge to domestic importers, licence holders get the quota rent. If the government auctions the licences freely, it captures the quota rent in the same way as it gets the tariff revenue in the case considered in Section 2*a*. Either way, the quota rent, represented by GHNS in Figure 1 (-assuming the quota to be set at GS on a nondiscriminatory basis), remains a part of A's gains from trade.

Consider now the formation of an FTA which frees B of the quota restriction but not C. It should be obvious that the results depend critically on how the quota on C is fixed. If the quota on C is fixed such that total imports from all sources do

---

[8] For further details, see Bhagwati and Panagariya (1996a) where the argument of this and the previous sub-section has been developed more fully. The imperfect substitutes model is discussed in detail in Panagariya (1996a and 1996b). Finally, for a summary of the theoretical literature, see Bhagwati and Panagariya (1996b).

© Blackwell Publishers Ltd 1996

not change after the union is formed, our previous analysis remains unchanged. In terms of Figure 1, the import quota on C is fixed at FS. Country B, though freed of the quota restriction, continues to export GF. No quota rent is generated on imports from B which now receives the full price, P_C^t. The results in column 3 of Table 1 are entirely unchanged.

A more plausible assumption is that the quota on C is fixed at the pre-FTA level of imports from that country. In this case, total imports into A will rise after the formation of the union, the price in A will fall and trade creation will take place. The net effect is still likely to be negative, however. For the partner's terms of trade will still improve (though by less than the full amount of the quota rent in the pre-FTA equilibrium) and a part of the rectangle GVUH will be lost. The basic conclusion of Section 2*a* remains valid.

3. NON-TRADITIONAL GAINS FROM NORTH-SOUTH INTEGRATION

The first regional movement which flourished in the 1950s and 1960s consisted of regional arrangements that were either North-North or South-South type. The original members of the European Economic Community and of the European Free Trade Area (EFTA) were all developed countries. Members of other regional schemes such as the Central American Common Market, Andean Pact, Association of South-East Asian Nations, and Economic Community of West African States were all developing countries.

The second regional movement, also referred to as the New Regionalism, can be traced to the beginning of negotiations for the Canada-USA Free Trade Agreement (CUFTA) in the mid 1980s. This movement began with the conversion of the United States to regionalism and has been characterised by several regional agreements between developed and developing countries. The most prominent of these is, of course, NAFTA which supersedes the CUFTA and is designed to create a Free Trade Area among Canada, Mexico and the USA. The second enlargement of the European Community in 1986 which gave Greece, Portugal and Spain entry into the Community also had elements of North-South integration. Finally, a large number of Association Agreements between the European Community (EC) on the one hand and Central and East European countries on the other represent examples of North-South agreements.

In principle, developing country participants of North-South arrangements can expect three advantages that are not available in South-South arrangements. First, such arrangements guarantee access to a large market. If the rich partner's markets are subject to trade barriers initially, the FTA helps to remove those barriers and, thus, *improves* the country's access to the former's market. If the rich partner's markets are already open, the FTA helps *maintain* the access in the event the former becomes more protectionist in the future.

© Blackwell Publishers Ltd 1996

Second, the FTA can shield the developing country member from administered protection in the rich country. For instance, the country may escape anti-dumping and safeguard actions by the rich partner to which other trading partners can be subject.

Finally, a regional arrangement with a large, rich trading partner can be an effective instrument of imparting credibility to reforms. An international treaty with a large and rich country can 'lock' the reforms, making it difficult for more protection-minded future governments to reverse the actions of their predecessors.

These benefits are not without qualification, however. First, to the extent that the southern country has higher initial tariffs, the FTA is likely to worsen its terms of trade. In the limit, if the northern country already has free trade, there is no change in the southern country's access to the former's market as a result of the FTA. On the contrary, it is the northern country which succeeds in acquiring access to the southern country *on a preferential* basis. In this case, the discriminatory access given by the developing country may be a hefty premium to pay in return for the promise by the rich country not to close its markets in the future.

Second, as was amply demonstrated during the NAFTA debate, integration with a southern country generates fears of adverse income distribution effects on unskilled labour in the North. These fears, in turn, lead to adoption of provisions (e.g., restrictive rules of origin) which partially undo the free-trade agreement in precisely those areas where it is likely to generate trade creation.

Finally, the benefits in terms of enhancing the credibility of reforms can be easily overstated. As far as tariffs are concerned, a country can attain the same objective *on a nondiscriminatory basis* through the GATT bindings. Regarding other instruments such as anti-dumping, an FTA does not offer any more restraint than the GATT. But that is likely to precisely enhance the possibility of such action against countries outside the union. This is the point made forcefully by Bhagwati (1993). To quote him:

> Imagine that the United States begins to eliminate (by outcompeting) an inefficient Mexican industry once the FTA goes into effect. Even though the most efficient producer is Taiwan, if the next efficient United States outcompetes the least efficient Mexico, that would be desirable trade creation ...
>
> But what would the Mexicans be likely to do? They would probably start AD actions against Taiwan ...

In the light of these benefits and costs of North-South regional arrangements, let us look closely at their evolution in the two major regions: North America and Europe.

© Blackwell Publishers Ltd 1996

4. NAFTA

In the Western Hemisphere, the North-South integration began with NAFTA. To assess the future desirability of similar FTAs, it is best to review briefly the main achievements of NAFTA. The following is the list of major changes to be implemented over a period of 15 years.

- Liberalisation of imports of *fresh fruits and vegetables* from Mexico into the United States and that of *corn* (and other grains) in the opposite direction.
- Liberalisation of tariffs and quotas on North American trade in *textiles and apparel.*
- Removal of most tariffs and non-tariff barriers (NTBs) on *cars* by Mexico within five to ten years.
- Mexico agrees to a rapid access for the US and Canadian firms to Mexico's *telecommunications* market. Mexico is to eliminate the majority of tariff and NTBs to its telecommunications equipment market upon implementation of NAFTA.
- In the *government procurement* area, over a period of 10 years, Mexico will open up to North American companies.
- NAFTA commits Mexico to a speedy implementation of the GATT *intellectual property* rights.
- Member countries agreed to provide national treatment to *investors* of another NAFTA member.
- The North American Agreement on *Environmental Cooperation* commits NAFTA members to improve environmental protection laws and enforce existing laws. Persistent failures are subject to a dispute settlement procedure which can assess monetary fines up to $20 million.
- The North American Agreement on *Labour Cooperation* does the same in the area of labour laws and labour standards.

Let us evaluate the likely effects of these changes.

a. Static Welfare Gains

From the viewpoint of static welfare effects, it is difficult to argue that Mexico will be a net beneficiary. Tariffs in Mexico are close to 20 per cent in many sectors affected by NAFTA while those in the USA and Canada are five per cent. Moreover, Mexico already enjoys preferential access to the United States market under the Generalized System of Preferences (GSP). According to Erzan and Yeats (1992), for the year 1986, Mexican exports to the United States which faced a tariff of five per cent or more and also had one or more significant nontariff barriers accounted for only 8.5 per cent of total exports. Hence, NAFTA is virtually certain to improve the US access to Mexico without a major change in

© Blackwell Publishers Ltd 1996

Mexico's access to the US market. Put differently, NAFTA will worsen Mexico's terms of trade.

The only sectors where Mexico appears to have won important new concessions *over other exporters* are fruits and vegetables and textiles and clothing. In the fruits and vegetables category, items in which Mexico had major interests such as orange juice and sugar are subject to special agreements that effectively limit import surges. And in textiles and clothing, Mexico's gains are likely to be limited for four reasons.

First, during the 1980s, Mexico consistently underutilised its Multifibre Arrangement (MFA) quotas with the United States. Erzan and Yeats (1992) place Mexico's average quota utilisation rate over 1985–89 in the United States at 51.5 per cent. Though, as discussed in Whalley (1995), quota underutilisation may mean many different things, this evidence suggests that the gain from the MFA-quota liberalisation may turn out to be rather limited. Second, under NAFTA, imports are subject to strict rules of origin based on 'triple transformation.' These rules are likely to limit Mexico's preferential access to the US market in textiles and clothing. Third, the gains from tariff liberalisation are limited due to a tiny share of textiles and clothing in total Mexican exports to the United States. The average tariff on Mexico's clothing and textiles exports in the United States in 1987 was 17.3 per cent.[9] The eventual removal of this tariff would seem to be a significant gain. But given that clothing and textiles account for less than two per cent of its total exports to the United States, the associated gain, even if such exports double, cannot be very large. Finally, to some degree, the dismantling of the MFA under the Uruguay Round will erode Mexico's preferential access to the United States' market.

By comparison, concessions granted by Mexico — virtually all of them on a preferential basis — are many and more substantial. Given the relatively high restrictions on outside countries — recall the highest tariff rate is 20 per cent — NAFTA is bound to lead to a substantial transfer of tariff revenue to the US exporters. The liberalisation in the automobile sector and government procurement is also likely to benefit mostly the United States. There may be some efficiency gains in the telecommunications sector where the US industry is said to be highly efficient but not enough to offset the losses in other areas. Whalley (1993) was right on the mark when, discussing the willingness of larger countries to go along with the demands of smaller countries for regional arrangements albeit at a price, he predicted:

> The current danger for Mexico in NAFTA is that the price will be paid in the form of adverse exclusionary arrangements which are against Mexican national interest, particularly autos, with potential impediments to third-country inward capital flows and arrangements which effectively reserve the Mexican import market for US suppliers.

---

[9] See Erzan and Yeats (1992), footnote 14.

© Blackwell Publishers Ltd 1996

b. Non-traditional Gains

In the long debate preceding the approval of NAFTA in the United States, economists were virtually unanimous that the arrangement was beneficial to Mexico. Not only the advocates of regionalism such as Dornbusch and Summers pushed for NAFTA, traditional sceptics such as Cooper (1993) and Corden (1993) also supported it. A key factor behind this consensus was the belief that NAFTA offered Mexico certain non-traditional gains not available through any other means.[10]

Surprisingly, the existence of these gains was accepted as a matter of faith. In all the writings prior to the approval of NAFTA by the United States — by believers and sceptics of regionalism alike — I have not found any critical assessment of these gains. Recently, Bhagwati (1995) has confronted the issue head on. Following him, let me attempt a full-scale critical examination of the non-traditional gains from NAFTA.

(i) Market access

According to this argument, NAFTA *improves* and *maintains* Mexico's access to the large North American market. I have already argued that with the exception of fruits and vegetables, *improvement* in Mexico's access to the North American market on account of NAFTA is minimal. This is not only because the US and Canadian markets are already quite open but also because NAFTA has 200 pages of rules of origin. Though the stated purpose of these rules is to ensure that goods do not get imported from a low-tariff country for transshipment to a high-tariff union member, given their complexity, they are bound to be misused for protectionist purposes.

An extension of the argument under consideration is that NAFTA guarantees Mexico *future* market access in case the US and Canadian markets close. In my view, this argument simply plays on the fears of weaker nations and has no basis in reality. I can imagine the United States threatening Japan, the EU, China, South Korea and even India and Brazil with new trade restrictions unless these latter opened their own markets further. But it is hard to imagine the use of such threats against a relatively open country such as Mexico with or without NAFTA. Given its own obligations under various GATT agreements, the United States is also unlikely to raise its own MFN tariffs. And in the unlikely event of an all-out trade war, irrespective of NAFTA, the United States is likely to move closer to, not away from, its trade partners to the south.

---

[10] Especially after Ross Perot declared his opposition to NAFTA, even economists critical of the regional approach found it difficult to oppose NAFTA lest they be viewed as supporting Perot's position and hence against free trade. The author is proud to say, however, that during December 1993 when the Congress was about to vote, he wrote and submitted to the *Wall Street Journal* an article entitled 'Death to NAFTA: Long Live Free Trade.' Of course, the Journal promptly rejected the article.

© Blackwell Publishers Ltd 1996

(ii) Shelter from new protectionism in the United States

It is sometimes suggested that NAFTA shelters Mexico from administered protection in the United States.[11] The reality, however, is that NAFTA provides Mexico with no more reprieve from anti-dumping actions than is available to other countries. Nor does it banish the use of safeguard actions. Indeed, the special agreements on sugar and orange juice explicitly allow for the play of administered protection in the event of import surges in these sectors. Furthermore, side agreements on labour standards and environment give the United States *new* powers to subject Mexico to dispute settlement procedures which can lead to fines of up to $20 million.

(iii) Credibility to and lock on the reforms

This is perhaps the most loudly expressed and least articulated argument made by proponents of NAFTA. As commonly expressed, the argument says that an international treaty with the world's richest country guarantees that reforms undertaken by the current government cannot be reversed by less reform-minded future governments. This, in turn, promotes a healthy environment for investment — both domestic and foreign — and growth. In its extreme form, the argument turns FTAs into a simple formula for growth: sign a regional pact with a rich country and you will prosper. The costs in terms of redistribution of tariff revenue to partner-country exporters, trade diversion and side agreements are then minuscule in relation to benefits. The argument definitely deserves a close examination.

The first point to note is that the causal relationship between NAFTA and credibility to reforms may be the reverse. Rather than NAFTA giving credibility to reforms, very likely, it was the (perceived) credibility of Mexico's reforms which led the United States to sign NAFTA in the first place. By the time the United States became serious about NAFTA, Mexico's reforms were perceived as irreversible even by markets. Without that perception, massive capital inflows including direct foreign investment into Mexico could not have taken place. Though the recent crisis has proved both the United States and the markets wrong, it does not detract from the fact that the perception of credibility played an important role in facilitating the negotiation, signing and passage of NAFTA. Put differently, if the United States Congress was voting on NAFTA today, without a shred of doubt, the vote would be negative.

We must also acknowledge that, ironically, NAFTA may have contributed to the recent macroeconomic crisis in Mexico. It is now becoming clear that the foundation of the crisis had been laid at least as early as 1993. Determined not to

---

[11] For instance, in a letter to the author, Jeffrey Schott of the Institute for International Economics writes, '... NAFTA is a GATT-plus accord that goes beyond the GATT commitment regarding contingent protection and import safeguard actions.'

© Blackwell Publishers Ltd 1996

let NAFTA slip away at any cost, Salinas' government used the massive capital inflows from the United States, caused to a large degree by recession and low interest rates there, to maintain an overvalued peso. This, in turn, allowed the Mexicans to maintain a high level of imports including consumer goods. A large proportion of these goods came from the United States. The imports promoted the image of a healthy Mexican economy on the one hand and increased the stakes of US exporters in Mexico on the other. Both factors helped to prop up support for NAFTA in the United States. But, in the meantime, much needed adjustment in the monetary policy and the exchange rate was delayed.

The crisis aside, can we still claim that NAFTA offers a lock on Mexico's reforms. To answer, we must be clear about what we mean by 'reform' in this context. If by reform, we mean *macroeconomic stability*, the recent crisis has already proved the argument to be false.[12] As a general point, macroeconomic stability comes largely from a country's own actions rather than the signing of an external pact. Chile does not have NAFTA but has managed its affairs well in recent years; it has had no crisis. Mexico had NAFTA but managed its affairs poorly; it suffered a crisis.

Next, let us look at microeconomic reforms. The most obvious first candidate in this area is trade policy. In evaluating NAFTA's contribution to *trade reform* and its credibility, we must compare it to tariff bindings offered by the WTO. Each member of the WTO has the option to bind its tariffs on different commodities at specified levels within the WTO framework. Once this is done, tariff increases beyond the bindings can be challenged in the WTO by countries whose exports are adversely affected. This is the mechanism which has sustained the tariff reductions of industrial countries under the various GATT rounds during the last four decades. Neither the United States nor Japan nor a host of European countries signed any regional pacts to lend credibility to their 'trade reform.'

There is no doubt that the WTO bindings offer a superior lock on tariffs than NAFTA. First, the WTO bindings are nondiscriminatory and apply to more than 120 countries while NAFTA locks tariffs on just two countries. In terms of economic efficiency, this makes the WTO mechanism superior. Second, on the face of it, violation of an agreement is less likely when there are dozens of potential challengers rather than just two. The old argument that compliance within the GATT/WTO framework is weak is no longer valid, now that a new, much tighter, dispute settlement mechanism is in place at the WTO. Finally, GATT bindings are not accompanied by rules of origin. Under NAFTA, the rules of origin have the potential to subvert the liberalisation even within the union.

---

[12] It may seem that proponents of NAFTA claim 'lock in' effects only on the policies actually covered in the agreement. But in policy discussions, I have found these claims to go well beyond the agreement. As an example, consider the following assertion by Jeffrey Schott in the letter to the author mentioned earlier, 'Furthermore, the "lock-in" effect applies to domestic policy reforms undertaken by the developing countries, and reduces the risk of policy reversals ...'

© Blackwell Publishers Ltd 1996

But, for the sake of argument, suppose for a moment that the proponents of NAFTA are right in that the NAFTA discipline works while GATT discipline does not. Now if protectionist pressures in Mexico rise in the future, how will the government respond? True, that the purpose of the lock is not to have to respond. But as we know from experience in virtually all countries, there are times when every government has to respond to protectionist demands. With NAFTA effectively locking the protection against the United States and Canada at zero and GATT discipline subject to abuse, the government may be left with no option but to raise tariffs against outside countries. With 70 per cent of Mexico's imports coming from the United States, this increase in the outside tariff will have to be very substantial indeed, to grant the desired level of protection.[13] By contrast, in the absence of NAFTA, the necessary increase in the tariff will be small and, being nondiscriminatory in nature, less damaging to Mexico's welfare.

Does NAFTA help to contain the use of nontariff barriers more effectively than GATT? As already noted, the agreement does not limit the use of instruments such as anti-dumping and safeguard actions which are governed by the WTO discipline. Nor does it provide any discipline against the use of direct import quotas against nonmembers beyond that provided by GATT. Some proponents of regionalism argue, however, that in times of macroeconomic crises, countries will not be able to resort to *exchange controls* if they are locked in a regional arrangement with a developed country.[14] As far as I know, this claim is incorrect. NAFTA creates an FTA among member countries but beyond that says nothing about the use of exchange or import controls in times of macroeconomic crises with respect to outside countries. Whatever discipline exists comes from the International Monetary Fund (IMF). Under Article IV, countries are required to consult with the IMF before they can introduce any new exchange controls. In the 1960s and 1970s, the Fund gave permission to use these restrictions with relative ease. But lately, the Fund has been more conservative. It should be remembered, however, that the Fund discipline which is weak in any case, applies to current account restrictions only. Countries, both developed and developing, employ exchange controls relating to capital account transactions.

Some proponents of NAFTA also argue that it will discourage the adoption of arbitrary domestic policies designed to favour one firm over others. This adherence to nondiscriminatory policies will, in turn, discourage rent-seeking activities. Once again, apart from the national treatment provision for investors of union members, there is nothing in the NAFTA document which supports this

---

[13] Since writing the original draft of this paper, after the recent peso crisis, Mexican tariffs on outside countries on 502 products have been raised to 35 per cent. Because pre-NAFTA tariffs in Mexico were 20 per cent or lower, this tariff increase implies a 75 per cent or larger tariff preference to NAFTA members.

[14] This point was made by Paul Collier of Oxford University at a World Bank Conference during a paper presentation by the author.

© Blackwell Publishers Ltd 1996

contention. But if national treatment itself is ill defined — note that the argument begins by asserting that the government discriminates between domestic firms — this provision cannot help very much. Perhaps the argument rests on a belief that the US firms likely to be hurt by the discriminatory policies will be looking over the Mexican government's shoulder and bring moral pressure on the latter to be even handed. Quite apart from issues of national sovereignty that this line of reasoning raises, its validity is highly suspect. If the US firms do acquire the ability to exert pressure on the Mexican government as a result of NAFTA, they will themselves seek preferential treatment. Domestic firms will, in turn, form alliances with US firms to continue their quest for rents. The net result will very likely be an increase rather than a decrease in rent-seeking activities and, worse yet, redistribution of rents from domestic to foreign firms.

Finally, proponents of NAFTA also argue that the provision on intellectual property rights will result in better copyright and patent laws and the side agreements on labour and environmental standards will lead to improved enforcement of the existing Mexican laws. Though I deal with this subject separately later, it may be noted here that these agreements were not sought by Mexico and, as such, not viewed by the country as being in its own interest. The view that higher standards are in Mexico's interest even if it did not seek them is not only paternalistic, it is also wrong. Higher labour and environmental standards in Mexico, without similar changes in competing developing countries, will surely lower Mexico's competitiveness in certain sectors. As for tighter intellectual property rights, at least in the short run, they will make technology imports more expensive. Reduced access to or increased prices of medicines are almost sure to hurt the poor.[15]

The contention of proponents of NAFTA that the arrangement will lead to a steady flow of foreign capital has been proved to be decisively wrong by the recent crisis. But for a Herculean international effort led by the United States — for which credit cannot go to NAFTA — Mexico was in danger of defaulting on its debt. As for direct foreign investment (DFI) rather than capital flows in general, the evidence provided by Whalley (1993) on the CUFTA should have a sobering effect on us. Whalley found that between 1987 and 1990, the DFI flow into Canada from the United States flattened while that from outside countries rose sharply.

5. A DIGRESSION TO THE EUROPEAN UNION

Lest I should appear opposed to *all* forms of regional integration, let me hasten

---

[15] See Panagariya (1996c) for a detailed analysis of the impact of tighter intellectual property rights on developing countries and Bhagwati (1996) and Bhagwati and Srinivasan (1996) for a discussion of the validity of the demands for higher labour and environmental standards by developed countries from developing countries.

© Blackwell Publishers Ltd 1996

to express my support for the European Union (EU) as defined by its core membership of 15 countries. To begin with, the origins of the EU were political and, regardless of whether it created or diverted trade, it yielded major political benefits. In its various incarnations, for several decades, the EU served as an effective bulwark against the Soviet bloc. For this reason alone, it was well worth having the EU.

But this is not all. On economic grounds as well, it is much easier to defend the EU than the NAFTA. As within the United States, and unlike the NAFTA, there is a strong commitment to deep integration within member states of the EU. The EU is intended to be a *Common Market* with a common external tariff and free capital mobility which it already has and free mobility of labour which it seeks. The common external tariff eliminates the need for the rules of origin. It also provides a genuine lock on trade reform: a member state cannot raise its outside tariff without all members being willing to do the same.[16] The EU also prohibits anti-dumping actions by member states against one another. Instead, members are subject to a common competition policy.

The strength of the EU's commitment to integration is also reflected in the existence of compensation mechanisms designed to help the members who either lose large amounts of tariff revenue or are poorer. Adjustment assistance and income convergence have been central goals of the EU. The Regional Development Fund was created in 1972 explicitly in response to the UK's fears that some of its regions would suffer on account of its entry into the European Economic Community (EEC). Funds were also earmarked in the mid-1980s to smooth the entry of Greece, Portugal and Spain.

Before turning to the Free Trade Area of the Americas, a word on the recent European Agreements (EAs) between the EC and the Central and East European (CEE) countries is in order. These agreements effectively create an FTA between the EU and each of the CEE members in six to ten years' time. Many of the criticisms of NAFTA apply to these agreements. What is different, however, is the political context and eventual expectations. Politically, the CEE countries see these agreements as an instrument of distancing themselves further from Russia — a way to lock-in their new-found freedom. Equally important, they also see the agreements as a first step towards full membership in the EU which can eventually turn the immediate economic losses into gains.

---

[16] This point is developed formally in Panagariya and Findlay (1995). We should not forget that the lock works in both directions: the common external tariff can impede liberalisation by weaker members if the dominant members in the union are more protectionist. This is clearly a risk for Uruguay who, as a member of MERCOSUR which is a customs union, must adhere to the latter's common external tariff. But the EU's record in this area has been excellent: it has not only lowered the common external tariff during successive GATT rounds but has also led new members such as Greece, Portugal and Spain into faster liberalisation than would have been the case otherwise.

© Blackwell Publishers Ltd 1996

6. THE FREE TRADE AREA OF THE AMERICAS

Let us now look at NAFTA from the viewpoint of Latin American countries other than Mexico. There are two main issues: how is NAFTA likely to impact them and should they make efforts to enter NAFTA. The first of these issues can be illustrated with the help of Costa Rica whose trade regime was studied carefully recently by the UNDP-World Bank sponsored Trade Expansion Program (1992). According to the study, fears have been expressed that NAFTA will lead to a displacement by Mexico of Costa Rica's exports to the USA and that it will also divert investments to Mexico. Because the United States alone accounts for more than 50 per cent of Costa Rica's exports and an equally high proportion of its imports, *prima facie*, such fears may seem justified. But as the TEP report points out, a serious injury can happen only in the case of products which are currently subject to high barriers in the United States and are important to Costa Rica. There are few products which satisfy both these conditions. For some exports into North America, common to Mexico and Costa Rica, the two countries both enjoy duty-free access. For example, this is true of the 'maquilla' electronics industry as well as Costa Rica's major traditional exports. Sugar is the major exception but in that case, Mexico's access has itself been limited by special agreements. Main areas where NAFTA can have some adverse impact is non-traditional agricultural exports which are not particularly large for Costa Rica and apparel. Not surprisingly, the TEP report concludes, 'Overall then, a Mexico-US agreement offers little reason for pessimism. . . .'

What applies to Costa Rica perhaps applies to a large number of other countries. As argued above, within NAFTA, there was limited room for discriminatory liberalisation in the United States and Canada. Discriminatory liberalisation by Mexico was more extensive but its market is relatively tiny. Not surprisingly, estimates of the effects of free trade within North America on third-country exports are generally small. Laird (1990) estimates that complete removal of tariffs within North America reduces the exports of other countries in the Western Hemisphere to the United States by less than 0.8 per cent. Erzan and Yeats (1992) estimate that the total trade diversion from tariff elimination in North America will amount to 0.5 per cent of US imports from non-members. Of this, only six per cent will fall on countries in the Western Hemisphere!

Turning to the issue of entry, I do not share the enthusiasm of many for a rush to sign an FTA with NAFTA and, eventually, create a Free Trade Area of the Americas. For reasons that must be clear by now, it is highly doubtful that Latin American countries stand to gain much from entering into an arrangement which gives their rich counterparts — the USA and Canada — a much greater preferential market access than the latter are able to offer in return. Table 2, taken from Lustig and Primo Braga (1994), provides the information on trade barriers in Latin America. Unilateral trade reforms have virtually eliminated quantitative

© Blackwell Publishers Ltd 1996

restrictions in the countries shown and lowered the average tariff rates considerably. Yet, the highest tariff rates at 20 per cent or more remain high. Discriminatory liberalisation with the United States under these circumstances is unlikely to confer major gains and will, very likely, lower welfare. Indeed, in the end, if FTAA does come into being, gains to developing countries in the region will come largely from improved access to each other's market than from any additional access to the US and Canadian markets.

A different twist is added to the issue if we bring in the possibility of the EU forming FTAs with countries in Latin America. Given the frequent claim that EU markets are less accessible to outsiders than North American markets, an argument can be made that Latin American countries may be better off seeking FTAs with the former. Not surprisingly, possibilities have been raised of an FTA between the Southern Cone Common Market (MERCOSUR) and the EU. Whether such an arrangement will be favoured by the United States, or not, and whether it can go ahead even if the latter is opposed, remains to be seen.[17]

If one goes by the intensity of trade, for a large number of countries in Latin America, the EC is a larger trading partner than the United States.[18] Tables 3 and 4 show the direction of exports and imports, respectively, of Mexico and ten other Latin American countries. Based on exports, for six of these countries — Argentina, Brazil, Chile, Paraguay, Peru and Uruguay — even the EC12 has been a larger trading partner in 1990 and 1993 than the United States.[19] Particularly striking is the relationship between MERCOSUR countries and the United States. The United States' exports to and imports from MERCOSUR have been 2.5 per cent or less of her total exports and imports, respectively. Proportions of exports of Argentina, Brazil and Uruguay to the United States show a persistent decline over the three years shown. Only Paraguay shows a slight increase in 1993 over the corresponding proportion in 1990. Another interesting case is Chile. The United States and the EC accounted for roughly equal proportions of Chile's exports in 1985 (28 per cent) but by 1993, their importance had declined. In 1993, the EC accounted for 25 per cent and the United States 16 per cent of Chile's exports. On the import side, the United States is more important for Chile than the EC. On balance, these data combined with more substantial restrictions in the EC lend some support to the view that many Latin American countries may be better off forming an FTA with the EC than with NAFTA.

If countries do rush to join NAFTA or the EU as they certainly will, it is critical that they minimise the margin of preference to the potential partner. The

---

[17] A referee has noted that one way to turn what is at present merely a declaration of intentions on the part of MERCOSUR and the EU into serious negotiations for an FTA will be for the United States to voice opposition to such a move. Thus, the US influence has its limits.

[18] As the analysis in Section 2 shows, the import share of the partner country is far from a compelling criterion to consider FTAs.

[19] Recall that the EC, now the European Union, has 15 members as of 1 January, 1995.

© Blackwell Publishers Ltd 1996

TABLE 2
Tariffs and Liberalising Reforms in Latin America and the Caribbean

Country (Pre-reform year, post-reform Year)[a]		Average Unweighted Legal Tariff Rates (%)		Tariff Range (%)		Coverage of QHs on Imports (% of Tariff, unless Otherwise Noted)	
		Pre-Reform	Post-Reform[2]	Pre-Reform	Post-Reform[2]	Pre-Reform	Post-Reform
Argentina	(1987, 1991)	42(p)	10	15–115(p)	0–22	62 (of dom. prod.)	8
Belize	1992		NA		5–65		NA
Bolivia	(1985, 1994)	35	7.5	0–100	5–10	NA	minimal
Brazil	(1987, 1993)	51	14.2	0–105	0–35	39	minimal
Chile	(1984, 1992)	35	11	35	11	minimal	0
Colombia	(1984, 1993)	61	12	0–220	5–20	99	1.4
Costa Rica	(1985, 1992)	53(p)	15(p)	0–1400(p)	5–20	NA	0
Dominican Republic	(1990, 1994)		NA		3–35		significant
Ecuador	(1989, 1994)	53	9,3	0–290	5–37	100	0
El Salvador	1992		NA		7–30		minimal
Guatemala	(1985, 1993)	50(p)	15(p)	5–90	5–20	6 (of dom. prod.)	0[1]
Honduras	(1985, 1992)	41(p)	15(w,p)	5–90	5–20	NA	0
Jamaica	(1981, 1993)	NA	20	NA	5–30	NA	0[1]
Mexico	(1985, 1994)	22.6	12.5(w)	0–100	0–20	28	1.6
Nicaragua	1992		NA		0–20		NA
Panama	1991	88.2	59.8		0–90		NA
Paraguay	(1988, 1991)	NA	16	NA	0–20(y)	NA	significant
Peru	(1990, 1994)	66	16	0–120	15–25	100	0[1]
Uruguay	(1987, 1993)	32	18(y)	10–55	6–20	0	0
Venezuela	(1989, 1992)	37	10(m)	0–135	0–20	40	10

Notes:

[a] where applicable.

(p) Including tariff surcharges; (w) production-weighted average tariff; (m) import-weighted average tariff; (y) for 1992.

[1] Some QRs exist for health and safety reasons

[2] In practice, several countries have higher actual tariffs than those which appear in the table. These are tariffs that have been applied to some specific goods and/or temporarily to control sudden import surges. See more in text.

Sources: Asad Alan and Sarath Rajapatirana, *Trade Policy Reform in Latin America and the Carbibbean in the 1980s*, Policy Research Working Papers (World Bank, 1993); ECLAC, Social Equity and Changing Production Patterns: An Integrated Approach (Libros de la CYPAL No. 32, 1992); Sebastian Edwards, 'Latin American Economic Integration', *The World Economy*, 16, 3 and staff reports (May 1993); GATT, *Trade Policy Review* (various issues); IDB staff reports, IMF, *Exchange Arrangements and Exchange Restrictions*, Annual Report 1993; World Bank, staff estimates and reports.
Taken from: Lustig and Primo Braga (1994).

© Blackwell Publishers Ltd 1996

TABLE 3
Exports from Country to Partner Region

Country	*World*	*Share*	*DCs*	*USA*	*EEC*	*DC Oth*	*LDCs*	*LAC*	*Asia*	*LDC Oth*	*Merco*	*Arg.*	*Braz.*	*Par.*	*ORO*
Exports 85															
Argentina	9,648	0.5	49.1	12.1	26.4	10.6	50.9	18.1	5.8	27.0	6.9		5.1	0.9	0.9
Brazil	28,182	1.5	70.0	28.9	28.2	12.8	30.0	8.6	7.8	13.6	3.2	2.2		0.6	0.4
Chile	4,139	0.2	70.8	20.7	32.2	17.9	29.2	14.5	10.5	4.1	7.9	2.1	5.5	0.1	0.2
Colombia	3,844	0.2	84.2	37.9	32.1	14.2	15.8	12.1	0.8	2.8	0.8	0.6	0.1	0.0	0.0
Costa Rica	1,220	0.1	79.1	46.7	17.3	15.1	20.9	16.9	0.3	3.6	0.2	0.2	0.0	0.0	0.0
Ecuador	3,493	0.2	66.7	56.5	6.1	4.1	33.3	8.7	23.0	1.7	0.5	0.3	0.2	0.0	0.0
Mexico	28,968	1.5	92.5	66.9	14.6	10.9	7.5	5.8	1.3	0.4	1.7	0.2	1.4	0.0	0.1
Paraguay	466	0.0	63.0	5.5	51.8	5.8	37.0	31.6	4.7	0.7	22.4	4.3	16.3		1.8
Peru	3,297	0.2	71.0	34.9	22.4	13.7	29.0	14.0	7.3	7.7	3.0	1.1	1.8	0.0	0.1
Uruguay	1,353	0.1	63.1	42.2	15.2	5.7	36.9	18.1	6.4	12.4	15.8	4.9	10.5	0.5	
Venezuela	13,914	0.7	84.1	49.1	20.9	14.1	15.9	14.3	1.3	0.3	2.0	0.0	2.0	0.0	0.0
United States	227,541	11.7	65.1		23.3		34.9	11.5	14.3	9.0	1.6	0.3	1.2	0.0	0.0
The World	1,943,220	100.0	70.9	18.6	34.3	18.1	29.1	3.8	11.3	14.0	1.0	0.2	0.7	0.0	0.0
Exports 90															
Argentina	13,844	0.4	52.6	12.0	31.9	8.7	47.4	25.0	7.7	14.7	13.7		10.9	1.2	1.6
Brazil	36,023	1.0	71.0	23.8	32.7	14.4	29.0	9.7	10.5	8.8	3.5	2.0		0.6	0.8
Chile	9,366	0.3	73.1	16.8	34.7	21.7	26.9	11.7	12.2	2.9	1.2	5.6	0.3	0.2	
Colombia	6,995	0.2	85.4	48.7	26.0	10.6	14.6	11.1	1.4	2.2	1.6	0.4	0.5	0.0	0.7
Costa Rica	2,156	0.1	87.6	51.3	25.3	11.0	12.4	10.5	0.8	1.1	0.3	0.0	0.3	0.0	0.0
Ecuador	3,003	0.1	76.2	51.5	14.4	10.3	23.8	19.1	2.2	2.6	0.8	0.6	0.2	0.0	0.0
Mexico	41,032	1.2	93.5	75.1	9.2	9.2	6.5	3.8	2.1	0.6	0.9	0.3	0.5	0.0	0.1
Paraguay	1,156	0.0	53.3	4.8	46.6	1.9	46.7	38.8	6.0	1.9	33.8	3.6	29.0		1.3
Peru	3,764	0.1	70.3	22.6	27.4	20.2	29.7	13.6	8.6	7.5	4.1	0.3	3.8	0.0	0.0
Uruguay	2,169	0.1	44.0	15.5	22.4	6.1	56.0	38.6	5.3	12.1	33.2	5.4	27.4	0.4	
Venezuela	17,895	0.5	77.2	55.5	11.5	10.1	22.8	20.3	2.0	0.6	2.3	0.0	2.2	0.0	0.0
United States	402,017	11.5	65.1		26.1		34.9	10.7	17.6	6.6	1.4	0.2	1.1	0.0	0.0
The World	3,498,250	100.0	73.4	14.8	40.4	18.2	26.6	3.2	13.3	10.0	0.8	0.1	0.6	0.0	0.0

© Blackwell Publishers Ltd 1996

TABLE 3
(Continued)

Country	World	Share	DCs	USA	EEC	DC Oth	LDCs	LAC	Asia	LDC Oth	Merco	Arg.	Braz.	Par.	ORO
Exports 93															
Argentina	14,059	0.4	42.9	9.2	27.2	6.4	57.1	37.4	6.7	13.1	23.9		18.7	1.8	3.4
Brazil	38,983	1.0	57.8	20.6	25.6	11.6	42.2	22.5	11.5	8.3	12.3	10.3		0.3	1.6
Chile	10,501	0.3	62.4	16.2	24.9	21.3	37.6	17.7	18.0	1.9	11.0	6.2	4.2	0.3	0.4
Colombia	7,824	0.2	73.1	41.8	22.3	9.0	26.9	24.2	0.8	1.9	1.7	0.9	0.8	0.0	0.0
Costa Rica	2,954	0.1	84.3	56.6	19.1	8.6	15.7	14.2	0.3	1.2	0.2	0.2	0.0	0.0	0.0
Ecuador	3,303	0.1	74.6	47.3	18.9	8.4	25.4	22.7	1.1	1.6	2.2	1.4	0.8	0.0	0.0
Mexico	51,274	1.3	93.5	79.5	5.8	8.2	6.5	4.8	1.5	0.1	1.1	0.5	0.6	0.0	0.1
Paraguay	843	0.0	45.9	6.4	34.6	4.9	54.1	52.2	1.7	0.2	40.2	7.1	32.3		0.8
Peru	3,372	0.1	63.5	23.7	26.1	13.7	36.5	18.7	14.8	2.9	4.6	0.7	3.7	0.1	0.1
Uruguay	1,940	0.1	38.3	14.3	20.1	3.8	61.7	49.7	7.2	4.9	41.1	17.9	22.2	0.9	
Venezuela	14,253	0.4	77.7	61.1	10.2	6.3	22.3	20.9	1.2	0.2	2.8	0.2	2.6	0.0	0.0
United States	487,491	12.8	56.4		21.4		43.6	16.4	19.3	7.9	2.2	0.9	1.2	0.1	0.0
The World	3,804,550	100.0	66.5	15.8	34.2	16.5	33.5	5.0	18.0	10.5	1.3	0.5	0.7	0.0	0.1

Source: Direction of Trade Statistics (IMF).

© Blackwell Publishers Ltd 1996

TABLE 4
Imports from Country to Partner Region

Country	*World*	*Share*	*DCs*	*USA*	*EEC*	*DC Oth*	*LDCs*	*LAC*	*Asia*	*LDC Oth*	*Merco*	*Arg.*	*Braz.*	*Par.*	*ORO*
Imports 85															
Argentina	3,813	0.2	64.2	18.9	31.2	14.0	35.8	31.8	1.5	2.5	16.4		14.4	0.4	1.7
Brazil	14,035	0.8	48.8	22.4	14.6	11.9	51.2	11.2	4.6	35.4	5.0	3.5		0.4	1.0
Chile	2,641	0.1	62.0	25.8	22.8	13.4	38.0	29.3	5.6	3.1	13.9	4.2	9.0	0.5	0.2
Colombia	3,895	0.2	76.1	37.7	20.4	18.1	23.9	21.6	0.8	1.5	6.3	3.4	2.6	0.0	0.2
Costa Rica	1,059	0.1	66.1	39.8	15.1	11.2	33.9	30.3	2.9	0.7	2.0	0.3	1.7	0.0	0.0
Ecuador	1,765	0.1	71.7	33.5	21.5	16.6	28.3	20.4	4.8	3.1	7.7	0.8	6.8	0.0	0.1
Mexico	18,432	1.0	95.0	74.0	11.4	9.6	5.0	3.4	0.9	0.8	2.6	1.4	1.2	0.0	0.0
Paraguay	660	0.0	35.8	15.0	14.3	6.5	64.2	58.7	4.5	1.0	57.6	11.0	45.7		0.9
Peru	1,625	0.1	72.0	30.5	23.7	17.8	28.0	24.4	1.6	2.0	15.8	10.0	5.6	0.0	0.1
Uruguay	760	0.0	33.8	8.5	19.0	6.4	66.2	38.2	1.6	26.3	32.4	13.0	18.5	0.8	
Venezuela	6,902	0.4	86.8	49.3	25.4	12.1	13.2	9.8	2.1	1.3	5.4	1.1	4.3	0.0	0.1
United States	327,540	17.7	65.0		20.0		35.0	12.3	16.5	6.2	2.5	0.3	2.1	0.0	0.0
The World	1,854,620	100.0	68.4	11.5	35.0	21.9	31.6	5.4	11.2	15.1	1.9	0.5	1.4	0.0	0.0
Imports 90															
Argentina	4,971	0.1	67.6	23.7	31.7	12.3	32.4	26.6	3.9	1.9	15.7		13.0	1.1	1.7
Brazil	28,945	0.8	43.0	17.5	16.2	9.3	57.0	12.5	1.9	42.6	7.7	4.9		1.1	1.7
Chile	6,670	0.2	63.0	25.1	22.4	15.6	37.0	24.4	7.7	4.8	14.9	6.9	7.3	0.4	0.2
Colombia	5,158	0.1	76.0	39.5	20.3	16.2	24.0	19.7	2.3	2.0	4.8	1.4	3.2	0.0	0.2
Costa Rica	2,019	0.1	68.2	49.1	10.0	9.0	31.8	25.6	5.7	0.6	3.2	0.6	2.6	0.0	0.0
Ecuador	1,923	0.1	69.8	35.4	21.5	12.9	30.2	25.0	2.7	2.5	8.8	2.2	6.6	0.0	0.0
Mexico	40,141	1.1	92.3	70.7	12.0	9.7	7.7	3.3	4.0	0.4	2.1	0.8	1.3	0.0	0.1
Paraguay	1,676	0.0	40.2	18.3	13.3	8.6	59.8	39.3	17.2	3.2	31.9	8.8	22.7		0.4
Peru	2,430	0.1	61.3	32.0	17.1	12.2	38.7	33.3	2.8	2.5	14.2	7.7	6.0	0.2	0.4
Uruguay	1,368	0.0	39.9	10.6	21.0	8.3	60.1	48.5	4.0	7.7	41.6	19.2	21.5	0.8	
Venezuela	7,435	0.2	82.2	41.8	29.1	11.3	17.8	13.5	2.5	1.8	5.6	1.9	3.6	0.1	0.0
United States	494,128	13.9	61.5		19.5		38.5	10.1	20.1	8.2	2.0	0.3	1.6	0.0	0.0
The World	3,556,480	100.0	68.8	11.1	38.4	19.3	31.2	3.6	12.7	14.9	1.3	0.3	0.9	0.0	0.0

© Blackwell Publishers Ltd 1996

TABLE 4
(Continued)

Country	*World*	*Share*	*DCs*	*USA*	*EEC*	*DC Oth*	*LDCs*	*LAC*	*Asia*	*LDC Oth*	*Merco*	*Arg.*	*Braz.*	*Par.*	*ORO*
Imports 93															
Argentina	16,191	0.4	56.3	23.3	24.4	8.6	43.7	31.5	10.4	1.8	24.9		22.6	0.3	2.0
Brazil	24,959	0.7	62.2	24.3	24.4	13.5	37.8	17.3	7.3	13.3	11.4	9.6		0.4	1.5
Chile	10,463	0.3	56.9	24.9	18.9	13.1	43.1	22.7	12.7	7.7	16.5	5.0	10.6	0.3	0.5
Colombia	8,680	0.2	71.0	37.2	17.6	16.2	29.0	24.6	2.4	2.0	6.7	2.2	4.3	0.0	0.2
Costa Rica	2,818	0.1	77.4	54.9	15.0	7.5	22.6	18.5	3.6	0.5	3.7	0.2	3.5	0.0	0.0
Ecuador	2,956	0.1	69.5	37.1	19.5	12.9	30.5	23.9	4.7	1.9	8.3	2.5	5.8	0.0	0.1
Mexico	59,868	1.6	91.5	69.5	11.6	10.4	8.5	3.4	4.8	0.2	2.1	0.4	1.7	0.0	0.1
Paraguay	2,605	0.1	36.4	20.0	7.9	8.5	63.6	50.9	11.6	1.2	48.2	10.8	36.9		0.6
Peru	3,649	0.1	54.1	29.3	13.4	11.3	45.9	39.7	4.9	1.4	15.4	6.7	7.5	0.2	1.0
Uruguay	2,683	0.1	43.3	9.4	27.5	6.3	56.7	49.4	5.0	2.3	45.5	16.4	28.9	0.3	
Venezuela	11,108	0.3	75.6	41.4	21.2	13.0	24.4	16.0	7.4	1.0	5.5	2.0	3.5	0.0	0.0
United States	581,848	15.5	58.4		16.9		41.6	11.7	24.0	5.9	1.6	0.2	1.4	0.0	0.0
The World	3,745,780	100.0	67.6	12.4	34.8	20.4	32.4	4.3	17.2	10.9	1.4	0.3	1.0	0.0	0.0

Source: Direction of Trade Statistics (IMF).

© Blackwell Publishers Ltd 1996

inescapable implication of this prescription is that they must liberalise externally as they liberalise within an FTA framework. For instance, a country such as India will be eventually tempted to seek entry into the EU. But before doing so, it is important for her to carry its unilateral trade liberalisation much farther. At the current tariffs, which go up to 50 per cent and average more than 35 per cent, the potential transfers of tariff revenue to EU exporters due to preferential access to India's markets can be very substantial.

One final point remains. If the analysis of this paper is correct, why is there such a rush on the part of the countries in Latin American to get into regional arrangements with each other and with the United States? An answer to this question is provided by Bhagwati (1995) as follows[20]

> If the question is raised because it is inconceivable that the governments of those countries would not be rational in their policy choices, then that assumption itself must be clearly rejected. For one thing, as we well know from aid experience and the economic development literature, one can seek something that *sounds* good but actually does harm. Besides, the objectives of the leaders may be diversified. Thus, for instance, they may expect to gain political kudos by going along with NAFTA because, by granting preferential access to US exporting interests, and through the implied underlining of Mexico's special relationship to the United States, they may gain the support of the United States in reaching out for prizes in a variety of unrelated political arenas. Thus, for example, in the absence of NAFTA and the willingness of then-President Carlos Salinas de Gortari to put almost everything on the line for its passage, can one seriously imagine that the United States would have gotten Mexico into the OECD or backed President Salinas for the important job of the director general of the WTO?

7. REGIONALISM AND HARMONISATION

Let me now turn briefly to the possible role of regionalism in promoting harmonisation of domestic policies in the member countries of a union. The interest in this area has been sparked primarily by the pursuit of harmonisation in areas of environmental standards, labour legislation, tax policy, exchange rates, monetary policy, product standards, competition policy, and intellectual property rights among countries within the EU and between the EU and countries with which it has Association Agreements. The side agreements concluded by NAFTA members on environmental and labour standards have also given rise to an interest in harmonisation issues.

Though space constraints do not permit a detailed analysis, it is worthwhile to make two points. First, as systematic analysis in various contributions in Bhagwati and Hudec (1996), particularly Bhagwati (1996) and Bhagwati and Srinivasan (1996) demonstrates, the case for harmonisation of policies across countries is too weak to deserve a serious consideration in either a regional or global context. Differences in labour and environmental standards are perfectly

---

[20] Footnotes 12 and 13, associated with the following quotation, have been dropped.

© Blackwell Publishers Ltd 1996

compatible with — indeed, under most circumstances, required for — economic efficiency.

Second, even if, for the sake of argument, we agree on the desirability of harmonisation of certain policies, the issue is tricky in the context of North-South regional arrangements. The pressure in such schemes is on the southern country to bring its standards up to those of the northern country. Because the optimal standard for the southern country is likely to be different from that for the northern country, such harmonisation ends up becoming an instrument of protection for the northern country. In addition, even when the southern country's standard is below the optimum, it is likely to conform to that in other southern countries. If it adopts the standards of the northern country, it will be placed at a disadvantage *vis-a-vis* its southern competitors. The inevitable conclusion is that if the actual standards in southern countries are below whatever is viewed as 'optimal,' the right institution for raising them is a multilateral, not bilateral, forum. Specifically, issues of intellectual property rights and high labour and environmental standards should be dealt with multilaterally by appropriate international organisations.

The inevitable implication of this discussion is that in considering harmonisation at a regional level, countries must satisfy themselves that the issue is primarily regional in nature. Once this is done, it is difficult to think of cases where developing countries can use the regional mechanism for harmonisation. Most of the examples I can think of are those of cooperation in infrastructure building including roads, dams and sharing of river water and have little to do with harmonisation of policies.

8. A SUMMARY OF CONCLUSIONS

This paper takes a negative view of NAFTA and FTAA. The main conclusions may be summarised as follows.

1. When trade is liberalised on a discriminatory basis as in regional arrangements, the mercantilist conclusion is essentially valid: A country gains from liberalisation by the partner and loses from its own liberalisation.

2. Because the United States had little tariff preferences to grant to Mexico while the reverse was not true, the static welfare effect of NAFTA is likely to be positive on the former and negative on the latter.

3. Non-traditional gains to Mexico from NAFTA were greatly exaggerated in the debate preceding its approval. NAFTA does not shelter Mexico from the US anti-dumping actions. The first systematic analysis of the 'lock-on-the-reforms' argument in this paper reveals that the role of NAFTA in landing credibility to Mexico's reforms is minimal, perhaps even negative. NAFTA locks at most of Mexico's tariffs against the United States and

© Blackwell Publishers Ltd 1996

Canada. But this is not necessarily a good thing; the welfare effect of this discriminatory lock is very likely negative and, moreover, if protectionist pressures rise in the future, it could lead to a substantial increase in discrimination against outside countries. As for the provision on intellectual property rights and side agreements on labour and environmental standards, they were not sought by Mexico but, rather, forced on it. Higher standards in Mexico in these areas without a corresponding change in other developing countries will make Mexico less competitive against the latter.

4. The EU and NAFTA are fundamentally different arrangements. The EU has had an important political dimension. For several decades, it served as a bulwark against the Soviet Union and may have even contributed to the latter's demise. On the economic front, the member countries of EU have a much stronger commitment to deep integration. The arrangement involves a customs union, has a free mobility of capital and seeks free mobility of labour. By contrast, NAFTA is a free trade area requiring complex rules of origin and is hostile to the movement of labour. Recent Association Agreements between the EU and Central and East European countries have many of the flaws of NAFTA. But these agreements have an important political dimension: they allow the Central and East European countries to further distance themselves from Russia. Moreover, to the extent that the agreements may lead to an eventual entry to the EU with full membership, in the long run, they may be desirable on purely economic grounds as well.

5. Static welfare effects of the proposed Free Trade Area of the Americas (FTAA) will very likely be negative on Latin America and positive on the United States. FTAA will do little to give Latin America a preferential access to the US market while the reverse is not true. Because the most efficient producers of many products are in East Asia, not North America, trade diversion will dominate FTAA.

6. MERCOSUR has more to gain (or less to lose) by joining the EU than NAFTA in an FTA. The member countries already have free access to the North American market. An FTA with the EU can help improve access to that market.

7. Great caution should be exercised in using regional arrangements to promote harmonisation of domestic policies. To begin with, harmonisation is not a desirable goal under most circumstances. Moreover, where harmonisation can be justified, its pursuit at a regional level can be counter productive. This is particularly true of North-South arrangements. Optimal standards in many areas (e.g., environment standards) are likely to be lower for a southern than a northern country. In cases where optimal standards are uniformly below the 'optimum' due to a race to the bottom, harmonisation must be effected on a global basis.

© Blackwell Publishers Ltd 1996

REFERENCES

Bhagwati, J. (1993), 'Regionalism and Multilateralism: An Overview', in J. de Melo and A. Panagariya (eds.).

Bhagwati, J. (1995), 'US Trade Policy: The Infatuation with Free Trade Areas', in J. Bhagwati and A.O. Krueger (eds.), *The Dangerous Drift to Preferential Trade Agreements* (Washington, DC: American Enterprise Institute for Public Policy Research).

Bhagwati, J. (1996), 'The Demand to Reduce Domestic Diversity Among Trading Nations', in J. Bhagwati and R. Hudec (eds.).

Bhagwati, J. and R. Hudec (1996) (eds.), *Fair Trade and Harmonisation Prerequisites for Free Trade?* (MIT Press: Cambridge, Mass., forthcoming).

Bhagwati, J. and A. Panagariya (1996a), 'Preferential Trading Areas and Multilateralism: Strangers, Friends or Foes?' in J. Bhagwati and A. Panagariya (eds.), *The Economics of Preferential Trade Agreements*, 1996 (AEI Press, Washington, DC, forthcoming).

Bhagwati, J. and A. Panagariya (1996b), 'The Theoretical Analyses of Preferential Trading Areas', paper presented at the American Economic Association Meetings, San Francisco, 5–7 January, 1996; to appear in the *American Economic Review. Papers and Proceedings* (May).

Bhagwati, J. and T.N. Srinivasan (1996), 'Trade and Environment: Does Environmental Diversity Detract from the Case for Free Trade?', in J. Bhagwati and R. Hudec (eds.).

Cooper, R. (1993), 'Round Table Discussion', in J. de Melo and A. Panagariya (eds.).

Corden, M. (1993), 'Round Table Discussion', in J. de Melo and A. Panagariya (eds.).

Dornbusch, R. (1993), 'Round Table Discussion', in J. de Melo and A. Panagariya (eds.).

Erzan, R. and A. Yeats (1992), 'Free Trade Agreements with the United States. What's In It for Latin America?' Working paper No. 827 (Washington, DC: The World Bank, January).

Krugman, P. (1991), 'The Move to Free Trade Zones', Symposium Sponsored by the Federal Reserve Bank of Kansas City, *Policy Implications of Trade and Currency Zones*.

Krugman, P. (1993), 'Regionalism versus Multilateralism: Analytical Notes', in J. de Melo and A. Panagariya (eds.).

Laird, S. (1990), 'US Trade Policy and Mexico: Simulations of Possible Trade Regime Changes' (mimeograph, Washington, DC: The World Bank).

Lipsey, R. (1958), *The Theory of Customs Unions: A General Equilibrium Analysis* (University of London, Ph.D. thesis).

Lustig, N. and C. Primo Braga (1994), 'The Future of Trade Policy in Latin America', in S. Weintraub (ed.), *Integrating the Americas: Shaping Future Trade Policy* (New Brunswick and London: Transaction Publishers).

Melo, J. de and A. Panagariya (1993), *New Dimensions in Regional Integration* (Cambridge, Great Britain: Cambridge University Press).

Panagariya, A. (1995), 'Rethinking the New Regionalism', presented at the World Bank Conference on Trade Expansion Program, 23–24 January, 1995.

Panagariya, A. 91996a), 'The Meade Model of Preferential Trading: History, Analytics and Policy Implications' (mimeo, Department of Economics, University of Maryland).

Panagariya, A. (1996b), 'Preferential Trading and the Myth of Natural Trading Partners', Working Paper No. 200 (Center for Japan-US Business and Economic Studies, Stern School of Business, New York University).

Panagariya, A. (1996c), 'Some Economic Aspects of the TRIPs Agreement' (mimeo., University of Maryland, College Park).

Panagariya, A. and R. Findlay (1995), 'A Political Economy Analysis of Free Trade Areas and Customs Unions', in R. Feenstra, D. Irwin and G. Grossman (eds.), *The Political Economy of Trade Reform* (Essays in Honor of Jagdish Bhagwati, MIT Press, forthcoming).

Primo Braga, C.A., R. Safadi and A. Yeats (1994), 'Regional Integration in the Western Hemisphere: Deja-vu All Over Again'?, paper presented at the conference 'New Investment Opportunities in Eastern Europe and Latin America' (Bethlehem, PA, Lehigh University).

Summers, L. (1991), 'Regionalism and the World Trading System', Symposium Sponsored by the Federal Reserve Bank of Kansas City, *Policy Implications of Trade and Currency Zones*.

© Blackwell Publishers Ltd 1996

Trade Expansion Program (1992), *Costa Rica. Strengthening Trade Links to the World Economy*, 101–105.

Viner, J. (1950), *The Customs Union Issue* (New York: Carnegie Endowment for International Peace).

Whalley, J. (1993), Regional Trade Arrangements in North America: CUSTA and NAFTA', in J. de Melo and A. Panagariya (eds.).

Whalley, J. (1995), 'The Impact of the MFA Phase Out on the Asian Economies', presented at the Asian Development Bank Conference on Emerging Global Trading Environment and Developing Asia (29–30 May).

Wonnacott, P. and M. Lutz (1989), 'Is There a Case for Free Trade Areas?' in J. Schott (ed.), *Free Trade Areas and US Trade Policy* (Washington, DC: Institute for International Economics), 59–84.

© Blackwell Publishers Ltd 1996

46

WHARTON REPROGRAPHICS

LIPPINCOTT LIBRARY
DOCUMENT DELIVERY SERVICE

East Asia and the New Regionalism in World Trade

Arvind Panagariya

1. INTRODUCTION

IN spite of the remarkable success of outward orientated policies for more than three decades, regionalism has reappeared on East Asia's policy agenda. Negotiations for the North American Free Trade Agreement (NAFTA), protracted talks at the Uruguay Round of the General Agreement for Tariffs and Trade (GATT), conversion of the European Community (EC) into European Union (EU), and the pursuit of aggressive unilateralism by the United States have led policy makers in East Asia to take another look at the regional alternative. In 1990, President Mahathir of Malaysia called for the formation of an East Asian Economic Group (EAEG) to counter the emerging NAFTA and the EC.[1] In January 1992, the Association of South-East Asian Nations (ASEAN) announced plans to form the ASEAN Free Trade Area (AFTA) by the year 2007.[2] Most recently, the Asia Pacific Economic Cooperation (APEC) which includes the United States has been actively focusing on issues of regional interest.

ARVIND PANAGARIYA is from the Department of Economics, University of Maryland, USA. He wishes to thank Ramgopal Agarwala, Vinod Thomas and two referees of this journal for helpful comments and to Sumana Dhar for excellent research assistance. The paper is based on the work done for the Regional Integration Initiative Study of the East Asia Region of the World Bank and is also a part of the World Bank project RPO 677-86. Findings, interpretations, and conclusions in this paper are entirely those of the author and should not be attributed in any manner to the World Bank, its affiliated organisations, or to members of its Board of Executive Directors or the countries they represent.

[1] The proposed membership included all the major players in East Asia. The United States vehemently opposed the idea. Because many of the proposed members failed to endorse the idea, the group never took off and was later recast as the East Asian Economic Caucus (EAEC).

[2] ASEAN was founded in 1967. The association has six members, Brunei, Indonesia, Malaysia, Philippines, Singapore, and Thailand. One of the objectives of ASEAN is to promote preferential trading among member nations. To-date, progress towards preferential trading has been minimal. Excluding Singapore which is a free-trading country, intra-ASEAN trade accounts for less than five per cent of the member countries' trade. Of this, less than five per cent is subject to any kind of preferential treatment.

© Basil Blackwell Ltd. 1994, 108 Cowley Road, Oxford OX4 1JF, UK
and 238 Main Street, Cambridge, MA 02142, USA.

WORLD ECONOMY 17 (6), Nov 94.

A central question emerging out of these developments is: Should East Asia follow the lead of the NAFTA and EU and pursue the regional route to sustain growth in the region; if yes, in what form and if not, why not?[3] In the present paper, I attempt to answer this question by examining the role of regionalism in East Asia at three levels. First, I evaluate the only serious attempts at regional integration in East Asia: the ASEAN. I argue that the costs of such sub-regional preferential trading schemes outweigh their benefits. Second, I evaluate the case for a region-wide, discriminatory bloc along the lines of the EU or NAFTA and conclude that though the *threat* of such a bloc may keep the NAFTA and EU from turning into closed blocs, its actual execution is a risky proposition. Finally, I evaluate the case for a simultaneous, MFN style, nondiscriminatory, region-wide liberalisation in East Asia. I argue that a case for such a regional approach can be made but it is unlikely to be pursued.

In Section 2, I provide an overview of East Asia with special reference to the region's trade flows. In Sections 3–5, respectively, I evaluate the three forms of integration: sub-regional groupings such as the ASEAN and AFTA, a discriminatory East Asian trading bloc and region-wide nondiscriminatory liberalisation. In Section 6, I offer brief concluding remarks.

2. EAST ASIA: AN OVERVIEW

In economic terms, there are ten major players in East Asia: the ASEAN-4 consisting of Indonesia, Malaysia, the Philippines, and Thailand; the Newly Industrialised Economies (NIEs) comprising Hong Kong, Republic of Korea, Singapore, and Taiwan (China); and China and Japan. In this paper, the term East Asia is used loosely to refer to these countries collectively. In the following, I provide a summary of East Asia's position in the global economy and a broad look at the members of the region. My objective is to create a simple but clear picture which the reader can carry in his mind while reading the rest of the paper. More detailed information is provided in a series of tables accompanying the text.

a. *East Asia in the Global Context*

During 1980s, growth rates in all major countries in the region, except the Philippines, exceeded the average growth rate in industrial market economies by

---

[3] It has been claimed by some (e.g. Krugman 1993) that a *de facto* trading bloc already exists in East Asia. If true, this will make the paper redundant. Fortunately, much of the systematic econometric evidence is to the contrary. Thus, see Frankel (1992) and Saxonhouse (1993).

© Basil Blackwell Ltd. 1994

TABLE 1
East Asia: Economic Indicators — Year 1990

	GNP Per Capita	*Share in World GDP (%)*	*Share in World Population (%)*	*Share in World Exports (%)*	*Share in World Imports (%)*	*Export to GDP Ratio (%)*	*Import to GDP Ratio (%)*
ASEAN-4	853	1.27	5.62	2.40	2.64	31	35
Indonesia	570	0.49	3.20	0.72	0.59	24	20
Malaysia	2320	0.20	0.32	0.82	0.79	69	69
Philippines	730	0.20	1.10	0.22	0.35	18	30
Thailand	1420	0.38	1.00	0.64	0.90	28	41
NIEs	6850	2.34	1.29	5.57	5.77	39	42
Singapore	11160	0.16	0.05	1.47	1.65	150	173
Hong Kong	11490	0.33	0.10	2.29	2.23	115	116
Korea	5400	1.12	0.77	1.81	1.89	27	29
Taiwan	7950	0.72	0.36	n.a.	n.a.	n.a.	n.a.
China	370	1.71	20.34	1.71	1.42	17	14
Japan	25430	13.54	2.22	8.02	6.37	10	8
East Asia	2630	18.86	29.47	19.10	17.37	17	16
East Asia (excl. Japan)	776	5.31	27.26	11.08	11.01	34	35

Sources: World Development Report, UN System of National Accounts and IFS Trade Statistics.

at least one per cent and by as much as 6.6 per cent.[4] East Asia's share in the world GDP, exports and imports grew steadily while its share in population declined slightly over the decade. In 1990, the region accounted for slightly less than one fifth of both the world GDP and world trade and one third of the world population (Table 1).

There have been three main changes in the distribution of East Asia's exports across major regions of the world during 1980–90. First, the share of less developed countries has declined by half. Second, North America's share increased sharply during the first half of the decade but declined in the second half. Third, as with North America and Western Europe, intra-regional trade has grown proportionately over the years. By 1990, the region had become the largest recipient of its own goods. Even excluding China for which data are not available for the entire decade, the region exported 32.3 per cent of the goods to itself, 31.9 per cent to North America and 20.7 per cent to Europe (Table 2).

b. *Economic Indicators of the Countries*

(i) *Per-capita incomes, growth, and centrality of Japan.* The countries in the region are at very different levels of development. With a per-capita income

---

[4] In making this comparison, we include Japan in East Asia though it is also included in industrial market economies.

© Basil Blackwell Ltd. 1994

TABLE 2
Direction of Exports — Year 1990

Exporter	*Partner*							
	North America	*Western Europe*	*Europe*	*East Asia (*)*	*Latin America*	*Africa*	*Middle East*	*South Asia*
North America	41.9	22.3	23.4	20.4	5.0	1.7	2.6	0.8
Western Europe	8.3	71.0	74.4	5.3	1.1	3.3	3.3	0.7
Europe	8.2	70.6	74.5	5.2	1.1	3.3	3.3	0.7
East Asia	31.9	19.8	20.7	32.3	1.9	1.6	3.0	1.5
Latin America	22.9	25.3	27.6	10.3	14.0	2.1	2.4	0.4
Africa	3.0	66.0	68.0	4.6	0.6	12.8	4.4	3.6
Middle East	17.8	48.6	53.0	9.1	1.2	3.6	8.5	0.9
South Asia	17.1	30.1	46.6	18.3	0.3	2.7	6.5	3.2

Note: (*) East Asia does not include China.

Source: UN COMTRADE Data.

of $25,430 in 1990, Japan is the richest country in the region (Table 1). Per-capita incomes in the remaining countries range from a low $370 in China to $11,160 in Singapore.[5] ASEAN-4 enjoy significantly lower per-capita incomes than the NIEs. Within ASEAN-4, Malaysia and Thailand are richer than Indonesia and the Philippines. Within NIEs, Singapore and Hong Kong enjoy higher per-capita incomes than Korea and Taiwan (China).

Economically, Japan is by far the largest nation in the region. Its share in the world GDP for the year 1990 was 13.5 per cent. Within East Asia, it accounted for approximately 70 per cent of the region's GDP. If we exclude Japan, East Asia's share in the world GDP drops from one fifth to one twentieth.[6] In terms of population, China accounts for 70 per cent of the region's population. If we exclude it, East Asia's share in the world population declines from one third to less than one tenth. The remaining economies in the region are small. With three exceptions — China, the republic of Korea, and Taiwan (China) — individual shares of the remaining six countries in the world GDP are less than 0.50 per cent. The combined share of China, the Republic of Korea, and Taiwan (China) is less than three per cent.

Real incomes have grown rapidly in East Asia during the past three decades. In the three NIEs for which information is available — Hong Kong, Korea and Singapore — real growth rates have ranged from six to ten per cent in each of the

---

[5] Per-capita incomes change dramatically if they are calculated at the purchasing power parity (PPP) rather than the nominal exchange rate. According to the recent IMF calculations, they rise almost by a factor of four in the case of China. In the case of Japan, they fall by a factor of 0.7.

[6] The picture is dramatically different, however, if we use GDP figures based on the purchasing power parity (PPP). According to the recent, PPP-based IMF figures, for the year 1990, China's share in the world GDP is six per cent compared with Japan's 7.6 per cent and Germany's 4.3 per cent.

© Basil Blackwell Ltd. 1994

three decades between 1960 and 1990. More broadly, with just three exceptions — Indonesia in the 1960s, and Japan and the Philippines in the 1980s — all ten economies in the region have grown at real rates exceeding five per cent per annum during each of the past three decades. With the exception of Botswana which has a population of 1.2 million only and has been blessed with regular discoveries of diamonds, no other country outside East Asia has been able to match these growth rates.

(ii) Importance of trade. Trade has been the undisputed engine of growth for virtually all economies in the region except perhaps Japan.[7] Export performance of the NIEs is well documented. Korea's export growth has been in the double digits during each of the past three decades. In the 1960s and 1970s, its exports grew at the phenomenal rates of 34 and 23 per cent, respectively. In the 1980s when world trade grew far less rapidly than in the previous two decades, exports of China, Korea, Malaysia, Taiwan (China) and Thailand still managed to grow at rates exceeding ten per cent.

Though trade regimes in East Asia can be characterised as generally open, there is a substantial variation in the level of protection across countries. Openness of Japan's trade regime has been a matter of great controversy. In terms of formal barriers, Japan's trade regime is as open as those of other developed countries. There are virtually no quantitative restrictions in the country and tariff rates are low. Yet, Japan's import-to-GDP ratio is well below those of other countries of comparable size and with similar income levels. This fact has led many economists to argue that Japan's markets are protected by informal trade barriers and are thus relatively closed to outsiders.

Among the remaining East Asian countries, Hong Kong and Singapore are textbook examples of free trading economies. For Korea and ASEAN-4 countries, Table 3 provides the relevant information. The most remarkable point to note is that quantitative restrictions in Korea, Malaysia, the Philippines, and Thailand have virtually disappeared. Tariff rates, though lower than the level in the early 1980s, remain high, especially in Thailand. Indonesia has high levels of tariff as well as quantitative restrictions. Effective rates of protection for manufacturing in Indonesia and Thailand exceed 50 per cent. China is also among the more protected economies of the region. The unweighted and trade-weighted mean tariff rates in China were 43 and 32 per cent, respectively, in 1992. The coverage of administrative import regulation extends to more than 50 per cent of the imports. But because China's import-GDP ratio has been rising rapidly and tariff revenue as a percentage of the c.i.f. value of imports is very low (5.6 per

---

[7] There is general agreement that Japan has relied on infant-industry protection in some cases. Its relatively low exports-to-GDP ratio at 12 per cent in 1980, 13 per cent in 1985 and 10 per cent in 1990 suggests that the momentum for growth must come predominantly from internal markets.

© Basil Blackwell Ltd. 1994

TABLE 3
Trade Protection in East Asia and the Pacific: Selected Economies

	Average Nominal Tariff (Unweighted) Per cent	*Per cent of Import Items Subject to Import Restrictions Per cent*	*Effective Protection Rate in Manufacturing Per cent*
Korea	10 (1992)	<5 (1992)	28 (1988)
Indonesia	20 (1991)	10 (1991)	52 (1992)
Malaysia	< 10 (1990)	<5 (1990)	23 (1988)
Philippines	24 (1992)	3 (1992)	32 (1992)
Thailand	44 (1988)	< (1991)	51 (1988)

Note: Effective Protection Rule (EPR) for Thailand excludes agro-processing and uses value-added at world prices as weights. Using the more standard weighting of value-added at domestic prices yields at EPR of 60.7 in 1988.

Source: World Bank.

cent in 1991), it has been suggested that China's regime is de facto much more open than is suggested by its formal barriers.[8]

Purely in terms of trade-to-GDP ratios, the smaller economies of the region are very open. As shown in the last two columns of Table 1, except in the case of China and Japan, exports and imports as a percentage of GDP are high. In 1990, ratios were in excess of 100 for Singapore. For ASEAN as a group, each of these ratios is bigger than 30 per cent and for the NIEs, almost 40 per cent.

(iii) *Direction of exports and imports*. Tables 4 and 5 show the direction of exports and imports of the countries for the year 1991. Beyond some special relationships to be noted below, the centre of gravity of intra-regional trade is Japan. In 1991, virtually all countries except Hong Kong and Singapore exported more than ten per cent of their goods to Japan.[9] The proportion of exports of the countries going to ASEAN-4 was quite small. With the exception of Singapore, these exports ranged from 2.7 per cent for Indonesia to 6.5 per cent for Taiwan (China). Exports going to China, potentially the second-largest market in the region after Japan, were also small.

There are two noteworthy subregional trade relationships within East Asia, both driven partially by inter-country investments. The first of these involves China, Hong Kong and Taiwan (China) and the second Indonesia, Malaysia and Singapore. As a part of the first relationship, both China and Taiwan (China) export a large proportion of their goods to Hong Kong. A large proportion of

---

[8] See World Bank (1994). We do not have information on Taiwan (China) but its trade regime is known to be generally open.

[9] Though the latest year for which Tables 5–8 provide information is 1991, in the text I rely on the 1990 UN Comtrade data. This is because the other latest trade data discussed in the paper are for that year and from the UN Comtrade data bases.

© Basil Blackwell Ltd. 1994

TABLE 4
Direction of Exports: NIEs, China, Japan and ASEAN-5 — Year 1991

Exporter	*Total Export (in billion $)†*	*NIEs*	*ASEAN-4*	*Partner*					
				China/ Singapore	*Japan*	*East Asia*	*EC*	*North America*	*Other*
NIEs	279.7	13.8	9.2	9.9	10.2	43.2	14.7	24.1	18.0
Hong Kong	76.0	11.5	4.4	35.2	2.7	53.8	13.8	15.1	17.3
Korea	68.3	8.9	6.2	0.4(3.1)*	17.9	33.0	14.5	28.9	19.8
Singapore	59.2	13.1	22.4	1.5	8.7	45.7	14.0	20.5	19.8
Taiwan (China)	76.2	13.1	6.5	(16.3)*	12.1	39.8	16.3	31.5	12.4
China	72.0	11.3	3.0	(5.5)*	19.8	34.1	23.5	29.0	13.4
Japan	314.9	21.3	5.1	2.7	0.0	29.1	18.9	31.6	20.4
ASEAN-4	100.6	23	4.0	12.9	23.0	52.3	16.1	19.4	12.2
Indonesia	29.1	21	2.7	8.3	37.0	64.8	12.8	12.6	9.8
Malaysia	34.4	33.7	5.6	23.3	15.9	57.1	14.8	17.7	10.4
Phillipines	8.8	12	4.3	2.6	20.0	37.7	18.6	37.3	6.4
Thailand	28.3	15.6	3.2	8.3	18.2	38.2	20.3	22.9	18.6
Singapore	59.2	13.1	22.4	(14.9)**	8.7	45.7	14.0	20.5	19.8

Notes:
• Figures in parenthesis are exports to Hong Kong.
•• Figures in parenthesis are exports by Singapore to Malaysia.
† Direction of Trade data adjusted for exports of China through Hong Kong.

TABLE 5
Direction of Imports: NIEs, China, Japan and ASEAN-5 — Year 1991

Importer	*Total Export (in billion $)†*	*NIEs*	*ASEAN-4*	*Country of Origin*					
				China/ Singapore	*Japan*	*East Asia*	*EC*	*North America*	*Other*
NIEs	282.9	11.8	8.6	2.2	25.1	47.7	13.3	19.4	19.6
Hong Kong	72.1	25.1	5.1	5.5	22.7	58.4	12.8	11.1	17.7
Korea	81.2	4.3	5.8	(0.9)*	26.4	36.5	15.5	25.5	22.5
Singapore	66.3	10.0	18.9	3.4	21.3	53.6	12.0	16.5	17.9
Taiwan (China)	63.3	8.2	5.5	(3.1)*	30.0	43.7	12.3	24.1	19.9
China	64.0	36.4	4.4	(27.3)*	15.8	56.6	13.3	15.1	15.0
Japan	236.6	11.5	11.3	6.0	0.0	22.8	13.5	25.9	37.8
ASEAN-4	112.2	20.7	3.9	9.7	26.3	50.9	14.5	14.9	19.7
Indonesia	25.9	18.2	3.0	6.6	24.5	45.7	18.2	14.3	21.8
Malaysia	36.7	25.8	4.2	15.5	26.1	56.1	13.6	16.1	14.2
Phillipines	12.9	19.9	5.3	3.7	19.4	44.6	10.2	21.5	23.7
Thailand	36.7	17.8	3.8	8.2	30.1	51.7	14.3	11.8	22.2
Singapore	66.3	10	18.9	(15.3)**	21.3	50.2	12.0	16.7	21.1

Notes:
* Figures in parenthesis are exports to Hong Kong.
** Figures in parenthesis are exports by Singapore to Malaysia.
† Direction of Trade data adjusted for exports of China through Hong Kong.

© Basil Blackwell Ltd. 1994

China's exports to Hong Kong is, in turn, re-exported to the United States and EU. The data in Table 4 have been adjusted to net out these re-exports. On a gross basis, more than 40 per cent of China's exports in 1991 went to Hong Kong. China also imported 27 per cent of its goods from Hong Kong.[10] Taiwan (China) exported 16.3 per cent of its goods to Hong Kong in 1991. As a part of the second sub-regional relationship of which the SIJORI growth triangle is a part, in 1991, Malaysia sold 23.3 per cent of its exports to and bought 15.5 per cent of its imports from Singapore. Indonesia sold 8.3 per cent of its exports to and bought 6.6 per cent of its imports from Singapore.[11]

Because of these sub-regional relationships, exports from Hong Kong, Malaysia and Singapore are more concentrated in East Asia than those from other countries. In 1991, Hong Kong, Malaysia, and Singapore, respectively, exported 53.8, 57.1 and 45.7 per cent of their goods to East Asia. At 64.8 per cent, the share of Indonesia's exports going to East Asia was also large but this was because of the large proportion of its exports going to Japan.

On the import side, the proportion of goods coming from within East Asia is large for virtually all countries. This is either due to a large proportion of imports coming from Japan or the sub-regional trade relationships mentioned above. In 1991, the proportion exceeded 40 per cent for all countries except Japan and the Republic of Korea (Table 5). For five out of the ten countries — China, Hong Kong, Malaysia, Singapore and Thailand — the proportion was in excess of 50 per cent.

Having summarised the relevant economic features of the major players in East Asia, we are now in a position to confront the central question of this paper: Should East Asia go regional? I will argue presently that in spite of the many *possible* gains, both economics and politics are against a discriminatory bloc in East Asia. Historically, East Asia has benefited greatly from an open world trading system. The region's future interests will be best served by a strategy which ensures that this feature of the world trading system does not change. A discriminatory bloc does not fit that bill.

Is there a role for an alternative form of regionalism in East Asia? A regional approach which encourages region-wide, nondiscriminatory trade liberalisation could still hold some promise. But because the terms-of-trade effects of such liberalisation are likely to be adverse, political feasibility of its implementation is

---

[10] For details, see World Bank (1994).

[11] SIJORI consists of Singapore, Johor, and Riau. The latter two are provinces of Malaysia and Indonesia, respectively. Though there is no trilateral agreement of any kind among the three governments, there are two formal documents signed by the governments of Indonesia and Singapore. These outline Singapore's role in the development of the Riau province and Indonesia's promise to supply water to and guarantee investments from Singapore. For more details, see Chia and Lee (1992).

© Basil Blackwell Ltd. 1994

in doubt. Moreover, the economic success of this approach is predicated upon the extent of trade expansion in Japan in the years to come. If a substantial expansion takes place, a simultaneous trade liberalisation by other major players in the region such as China, Republic of Korea, Thailand and Indonesia can lead to trade liberalisation around the world beyond the Uruguay Round. If trade expansion in Japan is limited, prospects for the success of this nondiscriminatory approach are also poor.

3. SUB-REGIONAL GROUPINGS: ASEAN and AFTA

I divide the discussion in this section into three sub-sections. Sections 3*a* and 3*b* describe the evolution of ASEAN in trade area and plans for the AFTA, respectively. Section 3*c* evaluates the case for promoting formal regional integration in the sub-region.

a. ASEAN: The Disappointing Past[12]

The ASEAN was founded in August 1967 with the intention of promoting economic cooperation in areas of trade, industry, energy, tourism , forestry, minerals, food and agriculture, and finance. In the area of trade, cooperation has been promoted through the Preferential Trading Arrangements (PTA) introduced in 1977. Additional programmes which have contributed to preferential trading include ASEAN Industrial Joint Ventures (AIJVS) and ASEAN Industrial Complementation (AIC) schemes. The discussion in this paper is limited, however, to the PTA.[13]

The PTA provides for tariff preferences, referred to as the margin of preference, for intra-ASEAN trade. Till the announcement of AFTA, the rules of origin required that the ASEAN content of a product be 50 per cent or more to qualify for tariff preference. On a case by case basis, this limit could be reduced to 35 per cent. Initially, items for tariff preferences were negotiated on a product by product basis. In 1980, an across the board minimum margin of preference was introduced for imports above a certain value but, because countries were allowed to have exclusion lists, the provision had little effect.

In 1987, preferences actually granted under the PTA were minimal. Based on the 50 per cent (or 35 per cent if agreeable) ASEAN content requirement, there were 12,783 items on the PTA list. Out of these eligible items, only 337 or 2.6

---

[12] For details on ASEAN until 1987, see Sopiee, See and Jin (1987). For more recent developments, see Pangestu, Soesastro and Ahmed (1992) and Ariff and Chye (1992). In the following, I draw freely on Pangestu et al.

[13] For details on the other two schemes, see the sources cited in the previous paragraph and Panagariya (1993).

© Basil Blackwell Ltd. 1994

per cent were actually granted tariff preferences. Furthermore, only 19 per cent of the total value of imports of these items enjoyed the preferential tariff.

At the Manila Summit in 1987, the member countries adopted changes aimed at strengthening tariff preferences. The countries agreed to shorten the exclusion list to ten per cent of the eligible items on the PTA list by 1992 (1994 in the case of Indonesia and the Philippines). They agreed to reduce the value of imports on the exclusion list to 50 per cent or less of intra-ASEAN trade by the same date. There was also an agreement to freeze the level of nontariff barriers and to negotiate reductions in them.

Systematic data on the progress towards achieving these goals is not available. But from what is available, progress appears to have been less than sparkling. Thus, in the case of Indonesia, the share of exports to ASEAN which benefited from tariff preferences rose from 1.4 per cent in 1987 to 3.5 per cent in 1989. Similarly, Indonesia's imports entering under preferential tariffs as a proportion of its total imports from other ASEAN countries rose from 1.2 per cent in 1987 to 1.6 per cent in 1989.

b. *AFTA: Time for Serious Business?*

Negotiations for the NAFTA and for a Single European Market in early 1990s swung the ASEAN members into action. At the Fourth ASEAN Summit in Singapore, on 28 January 1992, the member nations signed a framework agreement to establish and participate in the ASEAN Free Trade Area within 15 years, i.e., by the year 2007.

The key vehicle for implementation of the AFTA is the Common Effective Preferential Tariff (CEPT) which will be applied to goods originating from ASEAN Member States. The difference between the CERT and PTA is that the former is slightly more encompassing. The margin of preference under the PTA is granted only by the nominating country whereas that under the CEPT is granted by all members. The ASEAN content for qualifying for the CEPT is 40 per cent, lower than the PTA's 50 per cent.

The AFTA covers all manufactured products, including capital goods, processed agricultural products and those products falling outside the definition of agricultural products, in the CEPT scheme. Products for the CEPT scheme are identified on a 6-digit sectoral level of the Harmonised System (HS) of classification. Exceptions at the HS 8/9-digit level are permitted. The Third AFTA Council, held on 11 December 1992, identified a total of 38,680 items for inclusion in the CEPT. These items represent an average of 88 per cent of the total tariff lines of the ASEAN member states. The coverage ranges from 80 to 98 per cent among the six member countries. A total of 3,321 items have been identified for exclusion on a temporary basis. These items are to be reviewed at the end of eight years.

© Basil Blackwell Ltd. 1994

The Third AFTA Council also drew detailed schedules of reductions in CEPTs for all member countries. The Framework Agreement had identified 15 products as *Fast Track* products for a more speedy liberalisation.[14] For these products, tariffs above 20 per cent are to be reduced to 0−5 per cent by 1 January 2003, while those at or below 20 per cent are to be reduced to 0−5 per cent by 1 January 2000. The remaining products are on the *Normal Track*. Products in this category with tariff rates of 20 per cent or less are to be reduced to 0−5 per cent by January 2003 while those with tariff rates above 20 per cent are to be first reduced to 20 per cent by 2003 and then to 0−5 per cent by 2007. According to the schedules drawn by the Third AFTA Council, except Malaysia, no member planned any tariff reductions in the first two years of AFTA and even reductions planned by Malaysia were very small.

c. *Evaluating AFTA: A Wrong Turn*

As a forum for promoting political and cultural harmony and as an instrument of encouraging cooperation on projects of regional interest, the ASEAN has served the member countries well. But the net contribution of AFTA is likely to be negative. There are several reasons for this pessimism.

(i) Small internal markets. The case against the AFTA lies primarily in the small size of the regional market. Because Singapore is already a virtually free-trading country, the AFTA can, by definition, lead to no greater access to its market than what exists currently. Therefore, the gains from AFTA, if any, must come from integration of the remaining five countries' markets. But markets in these countries are quite small in relation to the world. The share of ASEAN-4 (the ASEAN minus Singapore and Brunei) in the world GDP at 1.3 per cent in 1990 is rather small. The region's share in the world exports is bigger — 2.4 per cent in 1990 — but not big. These small shares imply that, for a majority of products, the most efficient producers are likely to be located outside the region. Therefore, preferential trading will lead to a net trade diversion.

(ii) Low levels of intra-regional trade. An analysis of intra-regional trade flows tells us a similar story. The share of exports of Thailand going to ASEAN-4 has declined from 8.5 per cent in 1980 to 3.2 per cent in 1991 and that of Singapore has been in the range of 20−22 per cent. Shares of other countries, Indonesia, Malaysia, and the Philippines, show a rising trend but their levels are small. For three out of four ASEAN-4 countries, the share of exports going to

---

[14] These products include vegetable oils, cement, chemicals, pharmaceutical, fertilizer, plastics, rubber products, leather products, pulp, textiles, ceramic and glass products, gems and jewellery, copper cathodes, electronics, and wooden and rattan furniture.

© Basil Blackwell Ltd. 1994

Singapore has been generally larger than that going to the remaining ASEAN countries. The intensity of intra-regional imports, shown in Table 5, is equally discouraging from the viewpoint of promoting a free trade area in the region.

(iii) Uneven protection and uneven distribution of benefits of an FTA: Within the ASEAN-4, Indonesia and Thailand together accounted for approximately 70 per cent of the region's GDP and 56 per cent of its imports from anywhere in 1990. These are also the countries which are most protected within the region. Gains from discriminatory liberalisation, if any, must come primarily from liberalisation in these countries. But, in the first two years of AFTA, Indonesia and Thailand did not plan any *preferential* liberalisation.

Closely related to this point is the issue of the distribution of gains from preferential liberalisation when such liberalisation actually takes place. Preferential liberalisation is most likely to benefit Malaysia and Singapore at the expense of other countries. This is because much of the liberalisation will have to come from the more protected Indonesia, the Philippines, and Thailand. Recall that a very large proportion of imports of these countries comes from outside the ASEAN. Therefore, for many products, the most efficient producers are located outside the ASEAN. Discriminatory liberalisation under these circumstances will lead to trade diversion. While Singapore and Malaysia will benefit from a duty-free access to the markets in Indonesia, the Philippines and Thailand, these latter will bear the cost of trade diversion. The cost will come in the form of lost tariff revenue collected on the goods previously imported from countries outside ASEAN.[15]

These lopsided distributional effects may well explain why, to-date, the progress on *preferential* liberalisation in the region has been outstripped by unilateral, nondiscriminatory liberalisation. For example, while Indonesia and Philippines have lowered trade barriers substantially during the 1980s as a part of their trade reform policies, they have been generally reluctant to offer tariff preferences under the PTA. In July 1992, Indonesia announced a list of 250 tariff cuts but 90 per cent of these were on different types of batik cloth produced in Indonesia only. In the past, to lengthen their lists, member countries have gone so

---

[15] To make the point succinctly, let unit costs of production of VCRs be fixed at $200 in Korea, $240 in Singapore and $300 in Indonesia. Suppose that initially Indonesia imposes a $50 per VCR tariff on imports from all sources. All imports, say 10,000 VCRs in all, come from Korea at a tariff inclusive price of $250. Indonesia collects $500,000 in tariff revenue. An FTA between Indonesia and Singapore will allow Singapore to deliver VCRs at $240 which is below Korea's tariff inclusive price of $250. Indonesia will lose the tariff revenue. By contrast, if Indonesia is to liberalise imports on a nondiscriminatory basis, it will continue to import VCRs from Korea with consumers having to pay only $200 per VCR. Thus, under an FTA, the $500,000 tariff revenue is shared between Singapore producers and Indonesian consumers while under nondiscriminatory liberalisation, it is captured completely by Indonesian consumers. This analysis is modified under nonconstant costs of production but the basic message does not change.

© Basil Blackwell Ltd. 1994

far as to include snow ploughs among items to receive preferential tariffs! There have also been instances of tariff preferences on zero-tariff goods.

The distributional conflict is illustrated well by a remark made by the former foreign minister of Indonesia, Dr. M. Kusumaatmadja, at a meeting in 1992 to celebrate the 25th anniversary of the ASEAN:[16]

> Singapore and Malaysia are always telling us to lower tariffs and duties and let their goods into the country. But in return, how about the free movement of labour? We will take your goods if you will take our surplus labour supply. When they hear this and think about all those Indonesians coming to work in their countries, then they say, 'wait a minute, may be it's not such a good idea'.

(iv) *A comparison with NAFTA*. Though the inspiration to turn the PTA into AFTA has come, at least in part, from the EU and the NAFTA, it has little in common with the latter. Both EU and NAFTA are large markets. Therefore, from the viewpoint of securing markets and ensuring that the internal sources of supply are not much more costly than external sources, they can be defended to some degree. By contrast, the regional market of the AFTA is small and comes uncomfortably close to the markets in some of the regional associations in Latin America. Member countries, though richer than those in South Asia and Africa, are still classified as 'developing' and the scheme has the flavour of 'South-South' rather than 'South-North' or 'North-North' arrangements. The richest country in per-capita terms, Singapore, is too small in absolute terms to serve as the 'centre' country of the regional scheme as the United States is expected to do in the NAFTA. The inevitable conclusion is that an effective regional trading arrangement cannot be designed without the participation of Japan, an issue I turn to next.

4. AN EAST ASIAN TRADING BLOC? NO

There are two main issues which must be addressed in considering the case for a bloc consisting of all major players in East Asia: Is the bloc economically desirable and is the bloc feasible? I will argue below that the answer to the first question is at best uncertain while that to the second one is negative.

a. Is a Trading Bloc Economically Desirable?

The economic desirability of an East Asian bloc is difficult to assess. This is because the effects of such a bloc go well beyond simple efficiency effects. As shown in Table 1, the region accounts for approximately one fifth of the world's

---

[16] *Financial Times* (26 January 1993).

© Basil Blackwell Ltd. 1994

GDP and exports. Any major actions in the area of international trade at the region wide level which discriminate against the rest of the world will lead to repercussions in and perhaps retaliation from the rest of the world. Without being able to predict those reactions, it is difficult to estimate the costs and benefits of forming a region wide bloc.

The paramount objective of East Asia's regional trade policy has to be to ensure an open world trading system. Despite some redirection of trade towards itself in recent years, East Asia ships two thirds of its exports to the rest of the world. There is little doubt that the phenomenal growth of East Asia during the past three decades has been facilitated greatly by relatively open world markets. Almost without exception, studies of the NIEs draw a direct connection between growth in exports and that in the GDP. More recently, Indonesia, Malaysia, Thailand and China have been repeating the experience of the NIEs.

This suggests that the case for an East Asian bloc should be evaluated primarily, not on the basis of static gains including those arising from an improvement in the terms of trade, but in terms of its impact on the world trading system. If a regional approach is to be pursued, it should help keep the world markets open. There are two arguments in favour of a discriminatory regional bloc which deserve a close scrutiny.

First, an East Asian bloc may serve as a deterrent to the formation of closed trading blocs around the world. According to this argument, the world is already dividing into blocs. To ensure that the blocs do not become overly protective of their own markets and limit East Asia's access to them, East Asia should be united and be in a position to retaliate. Unilateral actions such as those taken by the United States under its Super 301 provisions will also be harder to take if East Asia is united.

Second, the Uruguay Round negotiations were protracted partly because of a large number of participants and the free-rider problem associated with it. According to this argument, larger countries feel that trade concessions negotiated by them become automatically available to smaller countries through the MFN clause of the GATT. Negotiations with smaller countries are difficult because their numbers are large and each of them is individually too small to make such negotiations worthwhile. Therefore, if the world can be first divided into a small number of blocs, it will be easier to organise future GATT negotiations.[17] Regional blocs would free up trade internally while the GATT process, once freed from the free-rider problem, can serve to bring down the barriers between blocs rapidly and with greater certainty. Proponents of the argument suggest that one reason why the past GATT rounds were so successful is that the United States could deal with the EU as a single unit. According to this

---

[17] The argument can be found in Summers (1991) and Krugman (1993).

© Basil Blackwell Ltd. 1994

view, if East Asia is turned into a bloc and the Americas into another, they together with the EU can move the world towards free trade faster.

Both of these arguments have some merit but are highly contentious. Regarding the first argument, critics note that countries organised into a bloc enjoy more market power than they do individually. Therefore, in principle, there is nothing to prevent blocs from *raising* rather than lowering trade barriers. The deterrence role of blocs is good only so long as the threat is not carried out. Once a threat is carried out and trade war breaks out, retaliatory actions are likely to be larger with than without blocs.

As for the second argument, critics note that small numbers do not necessarily mean faster progress.[18] The EU process began in 1957 and is still working towards a 'Single Market'. In the meantime, the EU's nontariff barriers have proliferated: the coverage of these trade restrictions has expanded five fold from 1966 and 1986.[19] It may also be argued that the Uruguay Round talks were not stalled because of the large number of players. The most serious disagreements in areas of agriculture and services were essentially two or three player game situations.

b Is a Trading Bloc Feasible?

Though the economic desirability of a trading bloc in East Asia is difficult to assess, its feasibility — or lack thereof — is more predictable. Both internal circumstances of the region and possible retaliatory actions from outside — particularly the United States — make the formation of an East Asian free trade area an unlikely event.

Internally, there are at least three inter-related factors at work against a region wide FTA. First, historically, the major players in the region have been political rivals. Though time, trade, and intra-regional investments have gone a long way towards bringing the former enemies closer, they still do not appear ready to form a free trade area with one another. In this respect, the situation in East Asia is fundamentally different than in Western Europe after the Second World War. Then, backed by the United States, for economic as well as geopolitical reasons, Europeans were able to move into treaties establishing first the European Coal and Steel Community and later the European Community. Today, there are no similar pressures on Japan. Nor are the other economically smaller nations such as Korea and China expressing eagerness to form an FTA with Japan.

Second, the countries in East Asia have very different levels of protection and are at very different stages of development. This makes the distribution of gains

---

[18] For example see Bhagwati (1993) and Winters (1993). A summary of the debate can be found in Melo and Panagariya (1992).

[19] See Winters (1993).

© Basil Blackwell Ltd. 1994

from an FTA rather uneven. With discriminatory liberalisation under an FTA, poorer countries which are also more highly protected are likely to lose or gain less than their relatively open and richer counterparts. This raises the spectre of compensation, a barrier not easily overcome.[20]

Third, the number of countries in the region is large which makes the task of far-reaching negotiations required for an FTA a daunting task. We saw earlier how difficult it has been for even six ASEAN countries to make progress towards the AFTA. It has taken the countries 25 years to reach the 'framework agreement' and progress on serious liberalisation is still out of sight. With this background, it is not clear how disparate countries such as China, Japan, Korea and the members of ASEAN can be engaged in a dialogue which will lead to a free trade area among them.

The external factors at work against an East Asian FTA are even more formidable. Because of perceptions that its markets are de facto closed to outsiders, Japan has been a persistent target of market-opening actions by the United States during the last two decades. These actions have included structural impediments initiative, and Super 301 threats. Smaller countries in the region such as Korea and China have also been subject to threats under Super 301. Initiatives by these countries for a free trade area, which can potentially divert trade from the United States, are almost certain to be met with retaliation by the latter.

From the viewpoint of smaller nations, this external environment is quite different from that faced by Mexico in negotiating the NAFTA. Apart from the fact that, with the United States as the other negotiating party the threat of overt retaliation did not exist. Mexico was simply not very vulnerable to such actions. In 1990, Mexico exported 71 per cent of its goods to North America and only 13 per cent to Western Europe and six per cent to East Asia. For the latter regions, imports from Mexico amounted to approximately a half per cent of their total imports.

The situation is dramatically different for countries such as China and Korea. They not only face an environment which is hostile to an FTA in East Asia but are also *individually* very vulnerable to actions against them by the United States. In 1990, Korea sold a quarter of its exports to the United States. China's direct exports were not as large but once re-exports through Hong Kong are taken into account, it too sent a quarter of its goods to the United States. With such large concentration of exports in the United States, risks for Korea and China of an FTA which the United States opposes are immense. This, in turn, suggests that an East Asian FTA is not a feasible proposition in the near future.

---

[20] The evidence on compensation schemes implemented in many regional arrangements in Africa is highly discouraging. See Foroutan (1993) for details.

© Basil Blackwell Ltd. 1994

5. NONDISCRIMINATORY, OPEN REGIONALISM? MAYBE

Having argued that sub-regional grouping and trading blocs which promote discriminatory liberalisation are not worth the effort required to create and sustain them, I now turn to a discussion of an 'open' regional approach centred around a GATT-style, MFN-based nondiscriminatory liberalisation. I argue that though this regionalism has certain advantages over discriminatory approaches, it, too, has serious limitations.

Before I consider the pros and cons of nondiscriminatory liberalisaton, let me take up briefly another type of open regionalism which has been suggested as a way to pre-empt the world from turning into inward-looking blocs. Bhagwati (1993), Cooper (1993), and others have suggested that an effective way to keep regional blocs from turning inward and ensuring that they eventually lead to multilateral free trade is to encourage them to have an open membership. Any country wishing to accept the obligations of the bloc membership should be allowed the privileges offered by it. Is such a bloc in East Asia feasible and desirable?

Without going into the desirability issue, let me simply note that the feasibility of an open but discriminatory bloc is highly questionable. Internally, one must confront all the problems noted in the previous sub-section. Externally, it may appear that the United States will be more tolerant of a bloc which is willing to accept it as a member. But the reality is otherwise. The United States is not ready yet to join in a free trade area with East Asia and without that, from its perspective, de facto there is no difference between an open and a closed bloc.[21]

This brings us to the remaining option: nondiscriminatory regionalism whereby countries in the region will pursue a GATT type liberalisation.[22] The key element distinguishing this approach from a regional bloc will be the nondiscriminatory nature of liberalisation. The countries in the region will come together at a common forum and, very much in the spirit of the various GATT rounds, negotiate reductions in trade barriers. Any concessions made by one country to another will be extended automatically to all GATT members. In the following, I offer a detailed discussion of the positive as well as the negative side of this approach.

a. Open Regionalism: The Positive Case

(i) No trade diversion. Because tariff reductions are nondiscriminatory

---

[21] The US position that it will neither join East Asia in an FTA nor allow the latter to form one on its own and also pursue NAFTA is obviously devoid of any internal logic. But this is the reality which East Asia cannot ignore.

[22] For a proposal along these lines, see World Bank (1993). Also see, Anderson (1991), Drysdale and Garnaut (1989 and 1993), Petri (1992) and the references therein.

© Basil Blackwell Ltd. 1994

under this approach, by definition, there can be no trade diversion. This lack of trade diversion takes away one major obstacle in the way of a discriminatory bloc: Countries with high initial tariffs need not feel that they will lose as a result of liberalisation. Any gains from liberalisation will accrue to consumers inside the country rather than the partner country. Problems mentioned earlier regarding compensation will simply not arise.

(ii) *External constraints*. This regionalism will certainly face a much less serious challenge from the United States than discriminatory approaches. Because the liberalisation is nondiscriminatory, it will improve the US access to East Asia's markets as well. The negotiations could go a long way towards answering the US complaints about a lack of openness of markets in East Asia in general, and Japan and China, in particular.

(iii) *Low adjustment costs*. Because liberalisation will take place simultaneously in all the major countries of the region, this approach will help to minimise the costs of adjustment. In the GATT style, liberalisation will be in areas of mutual interest. Therefore, countries will improve export prospects at the same time that they subject their import competing industries to competition from abroad. In contrast, if liberalisation is unilateral or, worse still, in response to Super 301 type of threats from the United States, adjustment costs will be higher. In the spirit of the GATT, it may also be possible to allow the liberalisation process to be spread over a period of, say, ten years. This will further smooth the path of adjustment.

(iv) *East Asia's role in world economic affairs*. In recent decades the United States has become what Jagdish Bhagwati calls a Diminished Giant. Simultaneously, East Asia, particularly Japan, has emerged as the major economic power in the world. Gradually, commensurate with their current economic weight and future potential, Japan and the Greater China region (including China, Hong Kong and Taiwan (China)) must assume the leadership role in world economic affairs. Japan has already emerged as a major donor country in the world. Within East Asia, it now enjoys the same central role as the United States in the Americas and Germany in the EC. It provided more than half of the official development assistance commitments to East Asia in 1992 against only six per cent from the United States. More than 30 per cent of the region's net direct foreign investment in 1990 came from Japan compared with ten per cent from the United States. Parallel with these developments in Japan, China is rapidly becoming a major engine of growth in Asia. According to a recent World Bank report, Greater China has become the world's 'fourth growth pole' after Europe, North America and Japan. The report notes that imports into the Greater China region are already two-thirds as much as Japan and will surpass the latter

© Basil Blackwell Ltd. 1994

by the year 2002 if growth continues at the present rate.[23] Import liberalisation by China in the years to come will further enhance that country's role in global economic affairs. Region wide liberalisation could then serve as a stepping stone to the eventual leadership role for East Asia in general and Japan and China in particular.

(v) Market driven integration. Nondiscriminatory liberalisation in the region may also reinforce the market driven integration which has been taking place in the region during the past decade. For example, liberalisation in China is sure to reinforce the trade and investment relationships in Greater South China. Likewise, liberalisation in Indonesia may help the growth process in the SIJORI region. To the extent that this liberalisation will reiterate loudly East Asia's commitment to outward orientation, direct foreign investment will flow into the region in increased volume.

b. *Open Regionalism: The Negative Case*

The discussion up to this point makes nondiscriminatory liberalisation almost too good to be true. And it is. Though, as just described, there is much to be gained from this type of liberalisation in the long run, the short to medium run economic effects are not favourable. This means that it will be difficult to mobilise support for implementation of the scheme. Let me elaborate.

The existing levels of tariffs, at least in Japan, are relatively low. This means that potential gains from lowering this most transparent barrier to trade in Japan are limited. Indeed matters are worse than revealed. Japan gives very extensive trade preferences under the GSP to its East Asian trading partners. For example, in the case of Korea, 88 per cent of the Japanese tariff lines facing it are either zero or below the MFN level. For approximately two thirds of the tariff lines, the GSP gives Korea a duty free access. A similar pattern applies to other countries.[24] This means that if Japan lowers it tariffs in a nondiscriminatory fashion, developing East Asian countries will lose the tariff preferences they currently enjoy. This is most likely to be a losing proposition for them.

There does not appear to be a substantial scope for gains from a reduction in tariffs in other countries either. Because tariff levels across countries are highly variable, the scope for quid pro quo is limited. Thailand, Indonesia and arguably China have the highest levels of tariffs. At the other extreme, Singapore and Hong Kong have virtually no tariffs while Japan imposes very low tariffs on its East Asian partners. In-between we have Korea and the Philippines with tariffs

---

[23] See Zuckerman (1993).

[24] Singapore pays a positive MFN tariff in only eight per cent of the cases. At 15 per cent, among the ASEAN countries, Indonesia has the largest share of tariff lines with positive MFN rates.

© Basil Blackwell Ltd. 1994

generally below 20 per cent. Given this cross-country structure of tariffs, it will be rather difficult to engineer an MFN style liberalisation.

The picture with respect to nontariff barriers is similar. Identifiable nontariff barriers are limited. We do not have information on all countries but in the cases where it is available, these barriers are not extensive. Japan employs few formal nontariff barriers. Among the countries shown in Table 3, Indonesia has the highest level of nontariff barriers. But even in this case, only 16 per cent of the country's import items are subject to such barriers. In terms of formal barriers, perhaps China is most protective. But in reality, China's imports have risen sharply in recent years suggesting that its import regime is freer than may be suggested by formal restrictions.

This discussion leads to the conclusion that the only significant trade liberalisation in the region may come from lowering the so-called 'informal' barriers to trade that Japan is often alleged to have. In order to bring to the negotiating table the countries with high formal trade barriers, only Japan can make attractive concessions. But the problem with 'informal' barriers is that either they may not exist or they are invisible. It is simply not clear how countries can negotiate on these barriers.

It is possible that purely on the basis of inter-temporal balance of trade, Japan's imports in the coming years will rise faster than its exports.[25] Then it may be argued that Japan can use this opportunity to aggressively lead other East Asian countries towards liberalisation. But once again, it is difficult to imagine other countries participating in a negotiation when import expansion in Japan is expected to happen through market forces in any case and there are no formal offers on the table for a reduction in trade barriers.

A final and important point is that even if we can somehow identify substantial trade barriers in Japan and other countries which can be negotiated away on a nondiscriminatory basis, the likely decline in the terms of trade of the region could be substantial. This is because more than 50 per cent of the region's imports come from outside countries. In the case of Japan, this number is as large as 70 per cent. Unilateral liberalisation by the region as a whole must worsen the terms of trade. Moreover, given that the level of protection in the rest of the world is not much lower than that in East Asia, it is very likely that such a decline in the terms of trade will be welfare worsening.[26]

This same problem — referred to as a free-rider problem — was faced by the major negotiating parties at the earlier GATT rounds. The United States, EC and

---

[25] The argument here is that Japan will not keep accumulating dollars for ever. It must spend them some time.

[26] Given positive barriers to trade in the rest of the world, the optimal external tariff by East Asia as a whole is perhaps higher than the tariff implied by the existing trade restrictions. Therefore, statically, unilateral liberalisation is likely to worsen the terms of trade as well as welfare.

© Basil Blackwell Ltd. 1994

Japan feared that benefits of liberlisation by them on the MFN basis will spill over to nonparticipating countries. But as Finger (1979) has shown, these countries were able to avoid this problem by limiting liberalisation to those goods in which they themselves traded intensively. Goods in which developing countries had a comparative advantage — agricultural products, textiles, shoes and other labour intensive products — were not liberalised. If East Asia is to follow this strategy, the scope for liberalisation will be limited. Given that approximately 70 per cent of Japan's exports and imports are with countries outside East Asia, a major trade expansion by that country will require liberalisation of products coming from outside the region.

In concluding this section, brief remarks may be made about the role of the East Asian Economic Group/Caucus (EAEG/C) and Asia Pacific Economic Cooperation (APEC). The EEAG was launched as a response to growing regionalism in North America and Europe and was intended to promote regionalism in East Asia on a wider scale than under AFTA. The proposed membership included ASEAN, NIEs, China and Japan. But US was opposed to the idea and was able to persuade Japan to stay out of the group. The 'group' never got off the ground and the concept was later recast as the EAEC which calls for period consultation among East Asia countries. At present, the EAEC appears to be dormant.

APEC was launched in 1989 with a wider membership than the EAEG. The membership includes the ASEAN, NIEs, China, Japan, Australia, New Zealand, Canada, USA and now Mexico. In 1992, a permanent APEC secretariat was established in Singapore. APEC's initial mission was cooperation in data collection and sectoral-policy analysis in areas of mutual interest and period consultation. In 1993, partially at the initiative of President Clinton, APEC came into prominence. APEC's meeting in December 1993 was attended by the heads of states of most member countries. Though APEC is now being viewed by its members as the natural forum for the discussion of issues of regional interest, given the diverse membership, it is unlikely to lead to major regional-integration initiatives.

6 CONCLUSIONS

On the whole, this paper takes a pessimistic view of regionalism in East Asia. The pessimism follows, inter alia, from low or negative gains in the case of sub-regional groupings such as the AFTA, insurmountable external and internal barriers to effective integration in the case of an Asia wide discriminatory bloc, and adverse terms of trade effects in the case of nondiscriminatory regionalism. In the ultimate, it regionalism is to be pursued, it is perhaps the last option which holds most promise. This option has the limitation that it may not offer large

© Basil Blackwell Ltd. 1994

opportunities for liberalisation and, if it does, the terms of trade effects on East Asia will be adverse. Yet, it has the advantage of promoting a more liberal and open world trading system. The long run gains from this openness far outweigh the short term losses arising from adverse terms of trade effects.

It bears repeating that the case against a discriminatory regional approach in East Asia is not based so much on its direct economic impact — either favourable or unfavourable — as on its infeasibility in the current environment. With many of the countries in the region selling an extremely high proportion of their exports to the United States, the ability of the latter to subvert a trading bloc in the region through both carrot and stick is beyond doubt. On the one hand, the United States can offer an FTA to countries such as Korea and Singapore in return for staying out of an East Asian bloc while on the other it can threaten them with retaliation if they go ahead with such a bloc.[27]

This reality combined with the persistent trade deficits of the US economy suggest that the best strategy for Japan and its East Asian trading partners today is to assume the same leadership role in the world economic affairs that England played in the 19th century and the United States did in the post-Second-World-War era. This means that, on the one hand, Japan should take a leadership role in pushing aggressively for continued openness in the world trading system beyond the Uruguay Round. It is likely that the emphasis on savings placed in Japan after the Second World War, which eventually led to the build up of huge trade surpluses, will now give way to increased expenditures and hence larger imports. A similar phenomenon is likely to play out in China. Taking advantage of this opportunity, Japan and China could take the high ground and lead East Asia and ultimately the world into a round of further trade liberalisation.

REFERENCES

Anderson, K. (1991), 'Is an Asia-Pacific Trade Bloc Next?' *Journal of World Trade* **25**, 27–40.

Ariff, M. and T.E. Chye, (1992), 'ASEAN-Pacific Trade Relation Relations', in *ASEAN Economic Bulletin*, **8**, 258–283.

Bhagwati, J. (1993), 'Regionalism and Multilateralism: An Overview', in J. Melo and A. Panagariya (eds.) *New Dimensions in Regional Integration* (Cambridge University Press, forthcoming).

Chia, S.Y. and Y. Lee (1992), 'Subregional Economic Zones: A New Motive Force in Asia-Pacific Development'. Paper presented at the 20th Pacific Trade and Development Conference, Washington, DC.

Cooper, R. (1993), 'Round Table Discussion', in J. Melo and A. Panagariya (eds.) *New Dimensions in Regional Integration* (Cambridge University Press, forthcoming).

---

[27] This is not to suggest that circumstances could not change ten years from now. For instance, ten years ago, it was difficult to imagine an FTA between the United States and Mexico but today it is more or less a reality.

© Basil Blackwell Ltd. 1994

Finger, J.M. (1979), 'Trade Liberalization: A Public Choice Perspective', in R.C. Amachen, G. Haberler and T. Willett (eds.), *Challenges to a Liberal International Economic Order* (Washington, DC: American Enterprise Institute).

Foroutan, F. (1993), 'Regional Integration in Sub-Saharan Africa: Past Experience and Future Prospects', in J. Melo and A. Panagariya (eds.) *New Dimensions in Regional Integration* (Cambridge University Press, forthcoming).

Drysdale, P. and R. Garnaut (1989), 'A pacific Free Trade Area?' in J. Schott (ed.) *Free Trade Areas and US Trade Policy* (Washington, DC: Institute for International Economics).

Drysdale, P. and R. Garnaut (1993), 'The Pacific: An Application of the General Theory of Economic Integration', in C.F. Bergsten and M. Noland (eds.), *Pacific Dynamism and the International Economic System*. (Washington, DC: Institute for International Economics).

Krugman, P. (1993), 'Regionalism versus Multilateralism: Analytical Notes', in J. Melo and A. Panagariay (eds). *New Dimension in Regional Integration* (Cambridge University Press, forthcoming).

Melo, J. and A. Panagariya (1992), *The New Regionalism in Trade Policy* (The World Bank, Washington, DC).

Panagariya, A. (1993), 'Should East Asia Go Regional? No, No, and Maybe' (World Bank, Policy Research Department, WPS 1209).

Pangestu, M., H. Soesastro and M. Ahmed (1992), 'A New Look at Intra-ASEAN Economic Cooperation', in *ASEAN Economic Bulletin*, 8, 3, 333–352.

Petri, P. (1992), 'One Bloc, Two Blocs or None? Political-Economic Factors in Pacific Trade Policy', in K. Okuizumi, K.E. Calder, and G.W. Gong (eds.), *The US-Japan Economic Relationship in East and Southeast Asia: A Policy Framework for Asia-Pacific Economic Cooperation*, Significant Issues Series, 14(1), (Washington, DC: Center for Strategic and International Studies).

Sopiee, N., C.L. See and L.S. Jin (eds.) (1987), *ASEAN at the Crossroads: Obstacles, Options and Opportunties* (ISIS, Kuala Lumpur).

Summers, L. (1991), 'Regionalism and the World Trading System', in *Policy Implications of Trade and Currency Zones* (Federal Reserve Board of Kansas City).

Winters, A. (1993), 'The European Community: A Case of Successful Integration?' in J. Melo and A. Panagariya (eds.), *New Dimensions in Regional Integration* (Cambridge University Press, forthcoming).

World Bank (1993), *East Asia and the Pacific Regional Development Review: Sustaining Rapid Development* (Washington, DC).

World Bank (1994), *China: Foreign Trade Reform. Meeting the Challenge of the 1990s* (Washington, DC).

Zuckerman, L. (1993), *Wall Street Journal* (17th May).

© Basil Blackwell Ltd. 1994

47

WHARTON/REPROGRAPHICS

Mercosur: Integration and Industrial Policy

Danny M. Leipziger, Claudio Frischtak, Homi J. Kharas and John F. Normand

I. INTRODUCTION

COMPRISING the majority of Latin America's population, land and economic output, the Southern Cone Common Market (Mercosur) represents the deepest attempt to date at economic integration in that region. Initiated by the Treaty of Asunción in 1991, Mercosur aims to eliminate all internal tariff and non-tariff barriers on the flow of goods and factors of production, implement a common external tariff, and harmonise numerous macroeconomic and sectoral policies among Argentina, Brazil, Paraguay and Uruguay. If it is successful, Mercosur members will gain from greater production rationalisation, more efficient resource allocation and expanded consumption opportunities.

More broadly, Mercosur is evolving into a core unit with which other countries and regional trading blocs can negotiate trade arrangements. Already agreements have been signed admitting Chile and Bolivia as associate members, and negotiations are taking place with the *Pacto Andino* countries and the European Union. Thus, Mercosur is engaged in simultaneous efforts to widen its membership and to deepen integration among its members.

The immensity and difficulty of this task is demonstrated by the fact that despite more than two dozen attempts at regional integration in the world over the past four decades, only the European Union has successfully forged the path from free trade area, to customs union, to common market, and now towards imminent monetary union.[1] The EU's experience has shown that the predicted welfare gains

---

DANNY M. LEIPZIGER is Lead Economist (World Bank). CLAUDIO FRISCHTAK is Economic Consultant (Brazil). HOMI J. KHARAS is Lead Economist (World Bank) and JOHN F. NORMAND is Researcher (World Bank). This paper is based in part on C. Frischtak, D. Leipziger and J. Normand, 'Industrial Policy in Mercosur: Issues and Lessons', World Bank Economic Note No. 10 (July 1996). All views expressed are those of the authors and do not represent the opinions or positions of the World Bank or its affiliates. Useful assistance was provided by Bryane Michael and Diane Bouvet. Comments were gratefully received from Julio Berlinsky.

[1] A free trade agreement (FTA) is one which removes all barriers to trade in goods among member countries while each member maintains its individual tariffs *vis-à-vis* non-member countries. A

© Blackwell Publishers Ltd 1997, 108 Cowley Road, Oxford OX4 1JF, UK
and 350 Main Street, Malden, MA 02148, USA.

from integration do not result automatically from tariff reduction alone, but rather depend on the removal of other barriers to competition.[2]

Indeed, one of the principal objectives of Mercosur is precisely to exploit potential welfare benefits from an economy-wide programme of reform and market liberalisation. By broadening the scope of liberalisation to an economy-wide free-trade area for goods and services, embedded in a lowering of common external tariffs, Mercosur explicitly seeks to deliver greater competition and efficiency through 'open regionalisation.' Closely tied to efforts to maximise welfare and promote open regionalism is the issue of industrial policy. Broadly defined, industrial policies encompass the combined effect of government policies designed to affect the allocation of resources among economic activities and produce an outcome different from what otherwise would have occurred in the market.[3]

Two tentative conclusions emerge from an analysis of Mercosur and its industrial policy: (i) the short-run welfare effects of Mercosur are likely to be small but positive, and (ii) industrial policy to date has not been very large but its future application must be judicious if Mercosur's goals are to be met.

This paper has two goals: to present the objectives and effects of Mercosur, as well as to discuss and analyse the Mercosur countries' industrial policies in the general framework of Mercosur. Section 2 sets out the objectives and the present status of Mercosur. Section 3 discusses some of the theoretical, policy, and empirical issues regarding Mercosur. Section 4 takes up the issue of when it is theoretically justifiable to use industrial policy, what the EC (now EU) experience has been, and what some major complementarities and conflicts of industrial policy are in Mercosur. Section 5 provides conclusions and recommendations.

2. MERCOSUR'S OBJECTIVES

Mercosur's main objectives are to establish a free-trade area with a common external tariff, and coordinated labour, monetary, tax, trade, R&D, regional development, and industrial policies. The Tariff Structure of Mercosur is shown in Table 1. To evaluate Mercosur's progress toward achieving these objectives, it

---

customs union is an FTA in which member countries adopt a common external tariff towards non-member countries. A common market provides for the free flow of goods and services, institutes a common external tariff and allows the free flow of factors of production such as capital and labour.

[2] In some cases these predicted gains can be quite large. Estimated static gains from the creation of the European Union range from two to six per cent of GDP. See Boltho (1982) and Molle (1990).

[3] Such policies could include tax exemptions or reductions for specific industries, selective credit schemes, concessional loans or administered trade protection measures such as anti-dumping measures and countervailing duties. See Leipziger and Petri (1993) for a review of industrial policies.

© Blackwell Publishers Ltd 1997

TABLE 1
Tariff Structure of Mercosur

Country	*Avg. Unweighted Tariff Rates (%)*	*Tariff Range (%)*	*MERCOSUR CET Avg. Rate (Unweighted)*	*CET Range*
Argentina[a]	12.3	0–20	12	0–20
Brazil[a]	12.6	0–35	12	0–20
Paraguay[b]	8.0	0–32	12	0–20
Uruguay[b]	14.7	0–20	12	0–20

[a] Source: World Bank. Brazil data 1996. Argentine data 1995.
[b] Source: Garay and Estevadeordal (1995).

is instructive to use the European Union as a gauge. Table 2 shows the objectives of Mercosur in comparison with the objectives of the European Union. To the extent that the European Union offers an example of a more 'evolved' union, it would provide a clear contrast which could be used to evaluate the progress and future design of Mercosur.[4]

Progress toward the first objective concerning the trade of goods has been steady. This is hardly surprising considering that the Mercosur treaty is the culmination of more than a decade of effort by Argentina and Brazil at pursuing difficult trade liberalisation. Major reorganisations in their trade regimes took place around the same time, with Brazil's 1988 tariff rationalisation (which was followed by a comprehensive trade reform) and Argentina's radical trade regime reorganisation in 1987. Principal reforms shared by both countries include the removal of most import barriers as well as a reduction in the level and dispersion of tariffs (shown in Table 1). In Brazil tariffs fell from 55 per cent in 1987 to 12.6 per cent in 1996 and in Argentina from 42 per cent in 1987 to 12.3 per cent in 1995 (in unweighted terms). The evidence points to the fact that the average tariff rate in Argentina is below that of Brazil when one deducts the three per cent temporary statistical tax imposed on most important categories. In any case, the actual unweighted averages are fairly close, both to one another and to the CET norm. In 1995, the common external tariff (CET) became effective bringing the tariff on most products between the range of 12 and 18 per cent.

It should be noted, however, that the CET currently covers only 85 per cent of harmonised tariff lines. The remaining 15 per cent of tariff lines (which cover almost half of all intra-Mercosur trade) are exempted temporarily to facilitate

---

[4] A caveat for this comparison is that Mercosur countries today differ considerably from the original EC members in 1960 in terms of level of development, industrial structure and regional disparities (all of which affect the nature of integration, the speed of implementation and the difficulty of policy harmonisation). The economies of the member states are more heterogeneous and not as interrelated as the European states. In terms of intra-regional integration, only 18 per cent of exports from Mercosur countries were sold within the bloc in 1994 (compared to an intra-regional export level of 33 per cent in the EC-6 at its inception in 1960), but up dramatically from 1990.

© Blackwell Publishers Ltd 1997

TABLE 2
Principle Provisions of Mercosur and the European Union

Policy Area	*Mercosur Objective*	*Status*	*European Union Objective*	*Status*
Trade of goods	100 per cent elimination of tariff and non-tariff barriers; implementation of common external tariff (CET); harmonisation of customs administration; establishment of rules of origin criteria.	■	100 per cent elimination of tariff and non-tariff barriers; harmonisation of customs administration; implementation of CET.	●
Capital mobility	Free movement of capital.	■	Free movement of capital; establishment of 'national treatment' laws for investment.	●
Labour mobility	Coordination of policies on labour relations, migration, training and safety.	○	Free movement of labour among member countries.	●
Monetary/ exchange rate policy	Coordination of monetary policies in order to reduce real exchange rate fluctuations and resulting disruptions in trade flows.	○	Coordination of economic policy and efforts to promote price and exchange rate stability.	■ ▼
Tax policy	Harmonisation of tax policies.	○	Harmonisation of sales, corporate and excise duty.	▼
Unfair trade practices	Establishment of common norms for administering anti-dumping measures and countervailing duties.	○	Establishment of judicial body to administer CVDs and anti-dumping measures	●
Dispute resolution	Establishment of dispute resolution mechanism and procedures to resolve trade controversies among Member countries.	○	Establishment of Court of Justice to resolve disputes among EU members; adoption of enforcement mechanism to ensure compliance	●
Decision-making body	Establishment of Common Market Group, Common Market Council and inter-governmental working groups in several policy areas.	●	Establishment of supra-national European Commision and European Parliament to consider legislation.	●
Research and development	Coordination of science and technology policies.	○	Promotion of EU-wide R&D initiatives in areas such as aerospace and electronics through specific budget budget outlays	●
Regional development	Harmonisation of regional development programmes.	○	Creation of funds (European Regional Development Fund and the European Cohesion Fund) to reduce regional development disparities.	●
Industrial policy	Harmonisation of industrial development policies.	○	Harmonisation/elimination of industrial promotion policies.	▼

Notes:

● Fully implemented	■ Partially implemented, but on shedule	▼ Partially implemented but delayed	○ Still under discussion

Source: Treaty of Asunción (1991); *General Report on the Activities of the European Communities*, EU Commission, various years.

© Blackwell Publishers Ltd 1997

adjustment but are all scheduled to converge to the CET within specified time limits (2001 for capital goods and 2006 for telecommunications and informatics, for example). The temporary exemption of these sectors implies that industrial policies are in place during the interim. Further, in terms of the other objectives, with the exception of capital mobility, one should notice that most of the objectives have only begun to be addressed. This implies that the major adjustment is yet to come.

3. MERCOSUR ISSUES AND PROGRESS

a. Trade Theory and Policy Issues

Economic theory provides strong underpinnings for the proposition that unilateral and multilateral liberalisation are the best policies for an individual or group of countries to follow, at least in terms of static economic comparative advantage. Unfortunately, theory is much less clear about the impact of regional trade arrangements (RTAs), in part because of the typically ambiguous results of second-best welfare economics, and in part because proponents of regional trade agreements in particular often point to dynamic effects which are difficult to capture.

From a static perspective, the impact of Regional Trade Agreements (RTAs) is usually analyzed by considering the effects of trade creation and trade diversion. Typically, it is argued that if trade creation dominates trade diversion, then the welfare of the *members* of an RTA may rise. However, it is inevitably the case that the welfare of *third-party* countries will fall. The crucial issue, then, is under what conditions will trade creation most likely outweigh trade diversion? Theory points to two factors. First, the higher the initial trade between two countries, the less likely it is that there will be sizable diversion due to regional preferences. It is presumed that initial trade already indicates global comparative advantage and so regional preferences do not result in switching purchases to less efficient producers. Second, the lower the external tariffs, the lower the likelihood of trade diversion.

In the case of Mercosur, these two effects do not yield clear-cut results. The initial trade between the two largest countries (Brazil and Argentina) was small. Common external tariffs, while significantly lower than in the past, remain at average levels of around 12 per cent (not negligible by global standards). As a further complication, a variety of sectoral exceptions, selected administrative interventions restricting internal free trade, and variable external tariffs add to the potential distortionary effects. Michaely (1996) is pessimistic. Looking at indicators of trade compatibility among countries, which rise with the level of development, the degree of industrial production, and the diversification of

© Blackwell Publishers Ltd 1997

exports, he concludes that there is little potential for trade creation gains from any regional trade schemes within Latin America. Yeats (1996) is even more negative, concluding that much of Mercosur's export growth has been in subsectors which remain the most protected from the rest of the world. In a similar vein, Garriga and Sanguinetti (1995) argue that most of the rapid trade growth within Mercosur is a consequence of aggregate trade liberalisation and should not be attributed to regional preferences. Their logic is that transport costs become more meaningful as tariffs and other trade barriers decline, and consequently neighbourhood trade should be expected to grow.

Nevertheless, most analysts conclude that trade creation effects have dominated trade diversion effects. Losada (1994) uses a partial equilibrium model of a variety of major sub-sectors and concludes that trade creation outweighs trade diversion for all countries except, possibly, Paraguay. One rough-and-ready rule of thumb which has been proposed for evaluating the impact on non-members is to compare the volume of imports into the region today with the volume before the Mercosur agreement went into effect. By this measure, it is clear that the joint process of multilateral and regional liberalisation that has occurred has been globally welfare enhancing. But even though the growth in intra-Mercosur trade has been rapid over recent years, the aggregate volume of intra-group trade remains below the level of what is considered a 'natural' trading bloc (Frankel, Wei and Stein, 1995). The implication of this market shallowness, combined with the inconclusiveness of opinion about the welfare effects of Mercosur suggest that while the welfare effects of Mercosur are probably positive, they are not very large at present.

While some analysts give much weight to the trade efficiency effects, others advance dynamic and political economy effects. The dynamic effects range from (i) a desire to exploit economies of scale that are not present in a single economy, (ii) the ability of a larger regional market to attract foreign direct investment, and (iii) a modified 'infant industry' suggestion that domestic corporations first need to be toughened through exposure to limited regional competition before facing the full force of global competition. In general, these arguments would have greater applicability to small countries. Mercosur, however, is driven by the two giants of Latin America — Argentina and Brazil — and consequently there is little reason to give much credibility to the dynamic arguments.

The political economy effects, perhaps more pertinent at this early phase of Mercosur's development, might affect these purely economic considerations.[5] The first is the issue of distributing costs and benefits. Economic differences, while not mitigating the potential economic benefits of integration, will determine the distribution costs and benefits and may engender political rivalries

---

[5] It is worth recalling that Mercosur issues largely fall under the Ministries of Foreign Affairs, and presenting common political objectives is expressly stated as a goal.

© Blackwell Publishers Ltd 1997

which work to frustrate integration attempts. Regional disparities among Mercosur countries (and for that matter, within countries) raise the likelihood that integration will aggravate regional income disparities in the same way that the accession of Portugal, Spain and Greece to the EU heightened the need for coordinating redistributional policies. Differences in the extent of economic liberalisation suggest that responsibility for policy harmonisation may fall disproportionately on some countries. A country like Argentina with relatively few industrial support programmes may find that harmonisation implies either forcing a faster pace of liberalisation on Brazil or 'ratcheting up' support to a level higher than the present. To what extent policy convergence will entail reversing the trend of liberalisation over the past five years is a critical issue for Argentine policymakers, who have to date been the fastest liberalisers.

The second issue is that of membership. The widening of group membership bolsters the objective of providing a coordinated body for negotiating with the other groups. Rapid expansion, therefore, can have its benefits. On the other hand, the larger the number of countries, the harder it becomes to harmonise rapid liberalisations. The tendency will be to move at a pace which all members can easily accommodate. However, provided that Mercosur continues to be associated with lower external barriers, the elimination of sectoral exemptions, membership expansion, and rapid growth of internal and external trade, it will be welfare enhancing and a stepping stone to greater openness in the region.

The third issue is how to specifically implement the treaty. Mercosur does not define any financing mechanism for regional development, industrial restructuring or reconversion initiatives, or joint research and development projects. Further, Mercosur's institutional arrangements are far weaker than those in the EU, which coordinate and enforce policy through supranational entities such as the European Parliament and the Court of Justice. Mercosur's inter-governmental working groups do not enjoy such legal autonomy or budgetary powers. Finally, the Treaty of Asunción's provisions for the harmonisation of monetary, tax and industrial policies do not specify either timetables or implementation mechanisms for achieving these goals. This ambiguity is neither unintentional nor necessarily undesirable at this stage; the range of policies to be harmonised and the number of constituencies involved is sufficiently broad and complex that member interests may be better served by specifying only the broadest objectives and allowing Mercosur's intergovernmental working groups to negotiate the terms of eventual policy harmonisation. However, Mercosur runs the risk of ambiguity, creating deadlocks due to disagreement, or derailment due to intransigence — dangers to Mercosur's objectives. In any case, the clear conclusion remains that much work still needs to be done.

The fourth issue is the extent to which adjustment will stimulate backpeddling. Already, Brazil introduced some non-tariff barriers and Argentina reinstated the

© Blackwell Publishers Ltd 1997

'statistical tax' which was really a three per cent tariff. Indeed, both Brazil and Argentina have found it increasingly difficult to grow rapidly without balance of payments pressures emerging. Both countries, although to different degrees, have eschewed the use of exchange rate policy to improve external competitiveness. Instead, they have attempted to bolster export growth by stressing the importance of reducing other costs to tradable goods producers through infrastructure development, financial system changes, and the alleviation of regulatory burdens. The result is that improvements in aggregate GDP growth can worsen trade deficits. While these pressures have been manageable to date, thanks to sizeable capital inflows, there is concern that such policies cannot be sustained over the medium-term. Once two key determinants of external capital flows (high domestic interest rates and significant privatisation opportunities) return to normal levels, the issue of trade deficits will become more critical. Both Brazil and Argentina have shown, in the recent past, a tendency to use trade interventions with the rest of the world and even with Mercosur partners as an instrument for managing balance of payments problems. But such solutions will become more and more problematic as Mercosur matures.

A final issue is how to deal with sectoral interests in the context of a customs union or common market. Many programmes that are important for national purposes, such as regional, sectoral, and small and medium enterprise development, social security and government procurement, have important implications for the location of foreign and local investment and trade. Recently, Brazil's decision to offer tax incentives for automobile production in its northeast region has been challenged by Argentina as an unfair trade practice, even though the decision was taken largely in response to the domestic issue of diverging regional development within Brazil. This example illustrates the hard choices that have yet to be made within Mercosur. As region-wide protection levels come down, the flexibility of the government to deal explicitly with hard domestic issues will also be reduced. Developing a new set of understandings as to how to handle these choices will be critical. Yet their unique problems, varying economic and political systems, and weak common institutional structures will likely complicate any mutually agreeable technical solutions. In this environment, political commitment to maintain the momentum of Mercosur will become even more important as a basis on which compromise solutions can be found.

b. The Results to Date

The contribution of general trade liberalisation and deeper liberalisation within Mercosur has led to rapid trade expansion. In 1988, all the Mercosur countries were relatively closed, with total trade/GDP averaging only 15.1 per cent. By 1995, this ratio had fallen slightly to 14.9 per cent despite trade expansion at a

© Blackwell Publishers Ltd 1997

TABLE 3
Mercosur Trade from 1988 to 1995

	1988			*1995*		
	*Intra-Mercosur**	*ROW*	*Total*	*Intra-Mercosur**	*ROW*	*Total*
Value (US$b):						
Exports	3.02	41.5	44.5	14.5	56.2	70.7
Imports	3.1	20.0	23.0	14.4	62.7	77.0
Total	6.1	61.5	67.6	28.8	118.9	147.7
Share of total (%):						
Exports	6.8	93.2	100.0	20.4	79.6	100.0
Imports	13.2	86.2	100.0	18.7	81.4	100.0
Total	9.0	91.0	100.0	19.5	80.5	100.0
Average Annual Growth (%):						
Exports	–	–	–	25.1	4.4	6.8
Imports	–	–	–	24.8	17.7	18.8
Total	–	–	–	24.9	9.9	11.8
Share of Mercosur GDP (%):						
Exports	0.7	9.3	9.9	1.5	5.7	7.2
Imports	0.7	4.5	5.1	1.5	6.3	7.8
Total	1.4	13.7	15.1	2.9	12.0	14.9

Notes:
* Totals for Intra-Mercosur exports and imports differ due to differences in reporting standards among trading partners. ROW refers to the rest of the world.

Sources: Trade data from IMF *Direction of Trade Statistics and Mercosur Journal.* Aggregate Mercosur GDP was obtained by summing the IMF's *International Financial Statistics* line 99b converted at the year's average exchange rate (IFS, line rf).

rate of 13.3 per cent per year.[6] Trade within Mercosur grew at 25 per cent, more than double the rate of trade with the rest of the world (10 per cent), but even in 1995 intra-Mercosur trade amounted to only one-fifth of the total trade of the bloc (Table 3).

In fact, the low level of total trade in Mercosur is still one of the principal characteristics of the region. Total imports of Mercosur countries are less than 8 per cent of GDP and total imports from other Mercosur countries are only 1.4 per cent of GDP. This compares with 4.5 per cent of GDP for intra-NAFTA imports and 14 per cent for intra-EU imports. Even if intra-Mercosur imports were to continue growing at 25 per cent until the end of the century, the regional imports would only reach 3 per cent of GDP. Thus, it must be understood that much of the expansion of trade and potential for future

---

[6] The 'openness' indicator of trade/GDP did not grow, as dollar GDP in Brazil and Argentina rose rapidly in keeping with significant real exchange rate appreciations following stabilisation programmes.

© Blackwell Publishers Ltd 1997

growth, both of intra-regional trade and trade with the rest of the world, will simply be 'catching-up' to international levels of openness for at least the next decade.

Trade volumes and values are at best indirect measures of the efficiency consequences of trade reform. Tybout (1992) and Rodrik (1993) suggest that three questions need to be asked at the microeconomic level: (i) has trade liberalisation reduced price-cost margins in the tradable sectors? (ii) has it resulted in scale economies or in other intra-industry rationalisation? and (iii) has it been associated with total factor productivity growth? Partial evidence is available on each of these questions.

There is little doubt that price-cost margins have come down in Brazil and Argentina since 1988. In Brazil, trade liberalisation was an explicit element of the stabilisation programme and tariff reductions beyond the original schedule were introduced in November 1994. In the ensuing two years, price increases for tradables have been consistently below those for non-tradables. For the period July 1994–June 1995, retail prices were 9.8 per cent higher than the average price index, industrial wages were 15.4 per cent higher, prices of industrial inputs were 16 per cent lower but prices of imported inputs were 29.2 per cent lower. This trend continues into June 1996, with machinery, cars and equipment prices 5 per cent lower than consumer goods prices, while health, construction and labour cost indices (as examples of nontradables) outpaced consumer prices by 36, 23 and 51 per cent, respectively. In Argentina, after the Convertibility Plan, the rate of growth of consumer prices far outstripped the growth rate of wholesale prices (by a cumulative 33 per cent between December 1991 to December 1994), before rising faster by 3 percentage points in 1995. And while costs did not necessarily increase to the full extent of general prices, being partly offset by tax reductions, export incentives and other reforms, most micro studies suggest that margins have been reduced.

Evidence on industry rationalisation is more anecdotal and sector specific. Nevertheless, there is some support for the thesis that intra-sectoral trade has been expanded, especially in the autoparts sector, and in some food processing sub-sectors. In Paraguay, substantial modernisation of the processed-meat sub-sector has taken place, with production becoming more oriented towards Brazilian markets. Argentina has found its exports of processed tomatoes and canned peaches falling substantially and being replaced by Brazilian producers able to access the same Argentine raw materials imports. These examples suggest some industry rationalisation. However, the interpretation is not clear-cut. The various exceptions to the Mercosur agreement permit substantial levels of effective protection to downstream industries, and have resulted in large intersectoral variations. Assessing industry rationalisation in this environment is fraught with pitfalls.

© Blackwell Publishers Ltd 1997

The evidence is more clear-cut with regard to total factor productivity growth. Among a group of Argentine sub-sectors studied by the World Bank (1996), strong TFP improvements were observed in dairy products, cellulose paste and plastics. Further, tradable goods, such as food and steel, had higher absolute competitiveness (relative to the United States) than non-tradable goods like banking and telecommunications. Also in Brazil, rapid productivity increases have been posted by industrial firms (approximately 40 per cent over the last five years) and surveys suggest much of this is attributable to TFP gains rather than to capital investment or increased use of intermediaries.

The evidence above addresses the broad question of the effects of regional and multilateral trade liberalisation. The more focused question about the impact of narrower liberalisation within Mercosur is harder to address and the evidence appears mixed. One point, however, seems clear: any welfare effects, whether positive or negative, are likely to be small in the short-run.

4. INDUSTRIAL POLICIES IN MERCOSUR

a. Conceptual and Historical Aspects of Industrial Policy

For several reasons industrial policy constitutes an obvious area for Mercosur's attention both during these early phases of trade liberalisation and throughout the entire integration process. First, selective industrial assistance programmes which distort relative prices will lead to resource misallocation and a loss of economic efficiency. Second, industrial policies introduce an additional complex political element since they provide a popular remedy for import-competing firms reluctant to adjust to trade liberalisation. Finally, integration may inspire competition among governments to outspend one another in order to confer a competitive advantage on domestic industries. Any positive benefit conferred by state aid in one country may be cancelled out by countervailing measures in another, leading to a squandering of resources. These reasons illustrate the fact that industrial policy is a powerful tool which could significantly affect the welfare outcomes of Mercosur. Thus, a common approach to industrial policy is necessary in order to (i) maximise potential gains from integration, (ii) provide clear rules of the game for governments in dealing with inevitable pressures for state aid in one form or another, and (iii) prevent actions, and reactions, which in the end may jeopardise Mercosur's success.

Broadly defined, industrial policies encompass any state acts or policies designed to affect the allocation of resources among economic activities and alter what would otherwise have been the market outcome. This definition is quite broad and can include sector-specific interventions, industry-specific interventions, or those applicable to all industries i.e., functional interventions.

© Blackwell Publishers Ltd 1997

Regardless of their national particularities, industrial policy instruments can be grouped under the rubric of trade, tax and credit policies.[7]

A number of reasons are cited for the use of industrial policy: to assist resource reallocation from declining to rising sectors, to correct externalities associated with specific industries, and to use 'strategic trade policy' (helping firms to capture an international pool of profits in globally oligopolistic markets). However, even though in theory these reasons might warrant intervention, faulty intervention often costs more than neutrality since poor targeting diverts resources from economically beneficial activities and distorts the allocation of inputs by warping price signals. It also protects poor industry decisions from suffering the full costs of their folly. Governments which follow a 'hunch' to support a particularly prestigious industry without clear evidence of a market failure, or those which identify favoured sectors based on political rather than economic criteria, will impose a high fiscal cost on the government and generally fail.[8]

The European experience is illustrative because it shows several trends which Mercosur might expect to emulate. First, the coordination of industrial policies with the goal of reducing state assistance to industry and minimising distortions in the common market has often been subordinated to political objectives. During the 1970s, the economic downturns induced by successive oil shocks encouraged member countries to grant increasing subsidies to industry in an attempt to preserve employment. Second, state aid to manufacturing in the EC from 1986 to 1990 fell by 16 per cent in constant terms. However, the overall aid level of $43b in 1990 still represented a sizable sum. While distortions should not be equated with the volume of assistance to manufacturing, its size and unsteady path of reduction should give Mercosur policy makers pause. Third, with the exception of Belgium, the poorer countries are heavier subsidisers. Spain, Italy and Ireland provided more than six per cent of manufacturing value-added in aid, which is 1.5 times the EC average. In Mercosur, one might expect Uruguay and Paraguay to subsidise their industry more heavily. Fourth, grants and tax reductions were the most popular industrial policy tools (possibly because they are more easily administered in terms of parliamentary procedures than other instruments).

These trends suggest a number of lessons to be learned from the European experience. Taking stock of the extent and nature of industrial policies on a

---

[7] Some examples of trade policies include selective duty rebates and non-uniform tariffs, non-tariff barriers ranging from outright QRs to more subtle protection such as product labelling and packaging requirements, safety standards, administered protection such as anti-dumping measures and countervailing duties, and differential rules of origin or domestic content requirements. Examples of tax policies include tax credits, allowances, exemptions or deferrals for particular firms, sectors, or regions. Some credit policies include below-market interest rate loans, government-backed loan guarantees, equity participation by the government and direct grants from the state budget.

[8] See Leipziger and Thomas (1993).

© Blackwell Publishers Ltd 1997

regional level is the first step in harmonising industrial policies. Second, it is necessary to establish strict guidelines for state assistance specifying the type of aid, absolute spending ceilings, and phase-out periods. Related to this rules-oriented approach to policy harmonisation, the EU experience emphasises the necessity of a credible institutional framework to monitor industrial subsidy programmes and enforce mutually agreed upon rules for industrial policy. Lastly, it is important to reconcile national policies and establish a supra-national policy framework. The first concerns the reconciliation of national policies to support domestic industry. This was the goal of the EC survey of state assistance programmes to industry initiated in 1989, the challenges of which have been noted. The second task involves the debate over the appropriate role for supra-national industrial policies to support bloc-wide industrial initiatives such as regional development programmes or downsizing of declining industries. Similarly, regional integration provides opportunities for collaboration on joint initiatives such as research and development in areas which were prohibitively expensive for a single country before integration. If national industrial policies in Mercosur are reduced only to be replaced by equally expensive regional initiatives, then the gains from more efficient resource allocation will be seriously compromised.

b. Some Industrial Policies[9]

To put Mercosur industrial policy in perspective, a review of industrial policies in Brazil and Argentina yields the overall conclusion that industrial policy interventions are on the whole relatively modest in both countries. Concerted industrial promotion efforts, as practiced in countries pursuing activist IPs (see Leipziger and Thomas, 1993, for a review) are largely absent, and industry-specific interventions, although they do exist, are limited in number. This is good news in terms of Mercosur harmonisation. Three examples of industrial policy in Mercosur are export promotion policies, regional development policies, and automotive policies.

Both Brazil and Argentina have been pursuing export promotion programmes which should boost long-run Mercosur-wide exports. The Brazilian export promotion programme consists of two key mechanisms: fiscal incentives (in the form of tax exemptions) and finance. Fiscal incentives come in five forms: tax exemptions, tax credits on domestic inputs for export-oriented production, a duty draw-back regime, tax concessions, and tax-exempt export processing zones. Direct finance for exports is available from the PROEX programme for a restricted positive list of goods and services (basically focused on capital goods

---

[9] More details on all these industrial trade programmes can be found in Frischtak, Leipziger and Normand (1996).

© Blackwell Publishers Ltd 1997

and engineering services). However, it should be noted that Brazil's export promotion policies do not seem to have been very effective. Between 1988 and 1993, the country's export-GDP ratio decreased from 10.4 per cent to 8.7 per cent, a little over a third of South Korea's. Brazilian exports as a percentage of world exports contracted in the same period from 1.2 to 1.0 per cent. The reason is because the domestic market is generally more profitable than export markets. Exports remain, for the most part, a countercyclical activity. Moreover, the relatively high cost of doing business in the country has become an effective impediment to the sustainable growth of exports.

In 1992, Argentina extended preferential duties to any industry group which agreed to set export targets. This plan, called the Industrial Specialisation Regime, was aimed at facilitating the productive restructuring of companies producing manufactured goods. Under the programme participating companies are required to increase their exports over those made the previous year in exchange for tariff preferences. The programme, although intended to meet the laudable goal of expanding exports, engenders several economic costs: foregone fiscal revenue which is estimated at $46 million through 1996, it subsidises an activity which may have occurred even in the absence of a government programme, it subsidises firms which should be willing to undertake the necessary restructuring and retooling in order to enter those export markets anyway, imposes a complex balancing requirement on imports and exports within a given sector in order for firms to qualify for special tariff exemptions, and embodies an implied discretionality in approving industries for participation in the regime and penalising firms which do not comply with their previously committed export quotas.

Both Brazil and Argentina have also been pursuing programmes which target the development of certain regions. In Brazil, federal tax incentives are only available for investments located in the Northeast and Amazon regions, and most states and many municipalities compete quite fiercely for new investments, through a mixture of explicit (sales and property) tax reductions and rebates, state banks' loans (although these are increasingly scarce), as well as through the supply of infrastructure services on a preferential basis. Although the fiscal cost of state and municipalities subsidies has not been precisely estimated, the amount of revenue foregone is a source of concern to the federal government in view of the fiscal condition of many states. In Argentina, legislation authorises provincial governments to provide incentives for regional economic development in the country's poorest provinces. Most programmes operate in San Luis, La Rioja, San Juan and Catamarca and offer one or a combination of the following incentives: exemption, reduction, suspension or deferment of national level taxes, accelerated depreciation, and duty free capital goods imports. In 1993 the fiscal cost of such programmes was estimated at approximately $1.54 billion. The establishment of free trade zones (with Tierra del Fuego being an important one)

© Blackwell Publishers Ltd 1997

also seeks to promote regional development. The Ministry of Interior estimated the value of these fiscal concessions at $226 million in 1994. However, these programmes have been criticised on the grounds that they have failed to achieve their stated goals of employment generation and industrial development, in addition to imposing a high annual fiscal cost. FIEL (1994), for example, notes that many projects approved for regional development would have proven unprofitable had they not received tax exemptions. Further, tax exemptions on capital goods have lowered the relative price of capital to labour and encouraged more capital intensive production, clearly against the goal of employment generation.

We have seen two examples of industrial policy in Mercosur, export promotion and regional development. The third example is that of one of Mercosur's most important industries, the automotive industry. The liberalisation of automobile imports in March 1990 had multiple impacts: it led to a fall in prices and the rapid expansion of domestic production, it brought about higher quality models allowed by more flexible domestic content rules, it stimulated competitive behaviour in an industry accustomed to quasi-cartelised conduct, and it attracted new entrants (including Renault, Mercedes-Benz, as well as Kia Motors and Asian Motors, among other South Korean automakers). Both assemblers and autopart producers are also directing their investments with a view of the common market, and setting up complementary production lines, in order to reap the considerable economies of scale (and product differentiation) which characterises the industry. Subsectoral trade is growing considerably, and reached US$ 1.14 billion in 1994 (including autoparts) — with Brazilian exports to the Mercosur reaching US$765.4b and Brazilian imports US$375.4b. In particular, the industry has been the major factor in the rising proportion of intra-industry trade between Argentina and Brazil. The intra-industry trade coefficient for the two countries reached 39 per cent of total trade for 1993, 51.3 per cent for manufacturing trade and 77.3 per cent in the case of the automotive industry.[10]

However, every silver lining has a cloud. Brazil's March 1990 trade reform led to an upsurge in imports at a rate the government felt would be unsustainable from a balance of payments perspective and highly damaging to domestic industry. So in June 1995, the government raised tariffs to 70 per cent and imposed quotas to deal with balance of payments pressures. As for Argentina, the automobile industry still heavily leans on the benefits conferred by the export-promoting 'industrial specialisation regime'. Without a doubt, the automobile

---

[10] See Lucangeli (1994). The intra-industry trade coefficient reflects the proportion of intra-industry trade within a particular segment relative to total trade volume in that segment. Thus, when the value of exports and imports within the industry are equal, the coefficient is 100 per cent. When positive exports (imports) correspond to no imports (exports) from within the industry, the coefficient is zero. Or algebraically, the intra-industry trade coefficient $\Pi = [1-([\text{abs. } (X-M)]/[X+M])]$ times 100 per cent.

© Blackwell Publishers Ltd 1997

sector is one which will continue to be fractious for Brazil and Argentina to handle in an advancing Mercosur.

The first-order estimate of the annual cost of industrial policies in Argentina, provided in Table 4, is about equal to one per cent of GNP, a modest amount. Initial estimates, which need to be refined, place Brazilian support for industry at around the same magnitude. At the national or federal level, therefore, the use of tax and credit policies to support industry in general, or certain portions thereof, such as the SME sector or exports, is small and generally visible. Temporary, and sometimes selective, tariff protection either in the guise of anti-dumping responses or reactions to balance of payments problems, has at times been used. However, the basic tendency has been to maintain openness.

While the scale of Mercosur industrial policy is still relatively modest, much work still needs to be done to rationalise Mercosur's industrial policies. To

TABLE 4
Argentina's and Brazil's Industrial Programmes and their Estimated Annual Cost in 1994

Argentina		*Brazil*	
Incentives	*US$(m)*	*Incentives*	*US$(m)*
Sector-specific		*Sector-specific*	
Quotas and special tariff concessions for auto sector	300	Automobiles	n.a.
Other sectors under Industrial Specialisation Regime	46	Informatics	297
Regional		*Regional*	
Tax concessions for regional economic development	1,500	Other	3,663
Tax concessions for investment in Tierra del Fuego	226		
Functional		*Functional*	
Concessionary tariff rates on capital goods imports	n.a.	Tax concessions for the technology development	350
Tax concessions for purchase of new capital goods	260	Tax concessions for purchase of new capital goods	646
Subsidised credit for SMEs and home mortgages	122	Subsidised credit for SMEs and home mortgages	534
Total	2,454		5,490

© Blackwell Publishers Ltd 1997

compare Mercosur policy with other regions of the world, the review of EC experience shows a very pronounced level of state assistance to manufacturing. While federal support of this magnitude in the Mercosur context should not be expected, there may be a concern that states in Brazil and provinces in Argentina will begin to vie for and subsidise industries in order to gain local employment. Some of this has already begun in Brazil with respect to large automobile plant investments. Experience from the EU shows that the problem of state-aid to industry is a pervasive one, and some early attention by Mercosur planners is warranted.

5. SUMMARY AND CONCLUSIONS

The two general conclusions of this paper are that the welfare effects of Mercosur, while not large in the short-run, are positive and that industrial policy interventions have been modest but still bear watching. The implication of these conclusions is that Mercosur will have to continue to evolve and progress in order to maximise its welfare-enhancing effects in the same way a dieter should redouble his efforts once the first pounds begin to melt so as to keep the momentum and not invalidate past efforts. Just like the European Union, Mercosur will have to face up to inevitable economic adjustment in certain sectors. This adjustment will be exacerbated by the fact that the member and associate countries are not as homogeneous as the EU was during the first legs of its union enterprise. Initial trade between its largest economies, Brazil and Argentina, is still small and at a level below what might qualify it as a 'natural' trading bloc. Further, while average tariffs in Argentina and Brazil have fallen from about 49 per cent to about 13 per cent (not negligible by international standards) and many economy-wide reforms have been taking place, several liberalisation measures will not take place until the end of the decade, and many specific policies still await agreement. Nevertheless, Mercosur has come a long way so far. It has in some measure created trade, lowered consumer prices, rationalised production, and promoted competitiveness *vis-à-vis* the rest of the world.

Because the range of gains from the Mercosur endeavour is both uncertain and highly dependent on dynamic effects, the role of industrial policy becomes crucial. If the major participants are prone to use IP interventions to protect declining industries and if national interests continue to prevail, then the road ahead will be a narrow one. Policy cooperation, concern for the consistency of policy, smooth sailing within the WTO, and resistance to ad hoc responses to deal with economic dislocations are of paramount importance. The reason for this is that so far policy interventions, while not always being optimal, have not been extremely large. Therefore, Mercosur could lose the fruits of its previous efforts or stymie future rewards if it is not diligent with regard to its industrial policy

© Blackwell Publishers Ltd 1997

oversight. Keeping this in mind, the following guidelines may be useful for Mercosur countries:

- *Transparency of industrial policy activities.* The failure to do this has often resulted in overlapping or competing incentives within the country, and a system which frustrates the comprehensive quantification of the costs and benefits of such policies.
- *Rapid phase-out of subsidy programmes.* Mercosur countries should set binding timelines for eliminating programmes which lack economic justification. Failure to do so for the EU has resulted in facilitating community-wide subsidisation rather than liberalisation, and in turn exacted a heavy resource cost.
- *Limit admissible types of intervention.* If some form of assistance to industry can be justified on sound economic grounds, governments should work towards adopting uniform policy instruments in order to promote greater transparency and reduce collateral distortions in other markets. Moving towards uniform instruments in this way can serve as a precursor to subsidy phase-out.
- *Consider only functional interventions.* Both theory and practice confirm the positive economic benefits of government assistance to industry in the case of market failures. Functional interventions, i.e., state aid directed at a general objective such as R&D without favouring a particular sector, can be welfare improving. However, government support of specific enterprises is generally recognised as costly for governments and seldom produces returns to justify its expense.
- *Anticipate requests for assistance from declining (sunset) industries.* If Mercosur is to work, resources will have to be reallocated and some industries will have to fold. Assistance programmes to facilitate adjustment should be temporary and conditional upon restructuring efforts such as reducing excess capacity.

There is considerable optimism for the future prospects of Mercosur. Mercosur 2000, the political programme designed to carry Mercosur into the 21st century, as well as discussions for a SAFTA (South American Free Trade Agreement) are underway. The Rose Garden Agreement signed in 1991 (setting up a forum for consultation between the US and Mercosur), and the Inter-Regional Agreement for Economic and Trade Co-operation between Mercosur and the European Union (which was signed by the Heads of Government in Madrid in December 1995) represent a growing recognition of the importance of the regional market in South America. As Mercosur finds itself involved in more multilateral and bilateral negotiations, momentum for Mercosur should be expected to build. If this is matched by caution in the area of industrial policy, Mercosur's potential can more easily be achieved.

© Blackwell Publishers Ltd 1997

REFERENCES

Boltho, A. (ed.) (1982), *The European Economy: Growth and Crisis* (Oxford: Oxford University Press).

Commission of the European Communities (1989), *First Survey on State Aid in the European Community in the Manufacturing and Certain Other Sectors* (Office for Official Publication of the European Communities).

Commission of the European Communities (1990), *Second Survey on State Aid in the European Community in the Manufacturing and Certain Other Sectors* (Office for Official Publication of the European Communities).

Commission of the European Communities (1992), *Third Survey on State Aid in the European Community in the Manufacturing and Certain Other Sectors* (Office for Official Publication of the European Communities).

Frankel, J.A., S-J. Wei and E. Stein (1995), 'APEC and Other Regional Arrangements in the Pacific', Pacific Basin Working Paper Series No. pb94–04 (Centre for Pacific Basin Monetary and Economic Studies, Federal Reserve Bank of San Francisco).

Frischtak, C., D. Leipziger and J. Normand (1996), 'Industrial Policy in Mercosur: Issues and Lessons', *Economic Notes* (World Bank, August).

FIEL (1994), 'Trade Policies Background Paper for Argentina: Country Economic Memorandum 1994' (November).

Garay, L.J. and A. Estevadeordal (1995), 'The Determinants of Regional Exchange in Mercosur: Geography and Trade Liberalisation', Universidad Torcuato di Tella, Working Paper No. 6 (Buenos Aires).

Garriga, M. and P. Sanguinetti (1995), 'The Determinants of Regional Exchange in Mercosur: Geography and Trade Liberalization,' Universidad Torcuato di Tella, Working Paper No. 6 (Buenos Aires).

Leipziger, D.M. and P. Petri (1993), *Korean Industrial Policy: Legacies of the Past and Directions for the Future*, World Bank Discussion Paper No. 197.

Leipziger, D. and V. Thomas (1993), 'The Lessons of East Asia: An Overview of Country Experience (World Bank).

Losada, F.J. (1994), 'Rubbing Out the Dividing Line: Trade Creation and Trade Diversion in the Common Market of the Southern Cone' (World Bank, mimeo).

Lucangeli, J. (1994), 'El Intercambio Argentino-Brasileno en 1994: Se Consolida un Padron de Comercio Maduro' (mimeo).

Michaely, M. (1996), 'Trade-Preferential Agreements in Latin America: An Ex-Ante Assessment', World Bank Policy Research Working Paper No. 1583.

Ministerio de Economia y Obras y Servicios Públicos de la Nación, Secretaria de Industria (1995), *Politica Industrial para una Argentina en Crecimiento*.

Molle, W. (1990), *The Economics of European Integration: Theory, Practice and Policy* (Worcester: Billings and Sons, Ltd).

Rodrik, D. (1993), 'Trade and Industrial Policy Reform in Developing Countries: A Review of Recent Theory and Evidence', in J. Behrman and T.N. Srinivasan (eds.), *Handbook of Development Economics* III (North Holland, Amsterdam).

Tybout, J.A. (1992), 'Linking Trade and Productivity', *The World Bank Economic Review* 6 (Washington).

World Bank (1996), 'Argentina: The Convertibility Plan: Assessment and Potential Prospects', Report no. 15402-AR.

World Bank (1996), 'Brazil: The Custo Since 1990–92', Report No. 15663-BR.

Yeats, A. (1996), 'Does Mercosur's Trade Performance Justify Concerns About the Effects of Regional Trading Arrangements?' *World Bank Policy Research Working Paper Series* (mimeo).

© Blackwell Publishers Ltd 1997

The Worl…
Vol 20 #4

48

LIPPINCOTT LIBRARY
DOCUMENT DELIVERY

WHARTON REPROGRAPHICS

NAFTA Expansion: US Imports Upon Chilean, Andean Pact, and MERCOSUR Accession

Michael A. Anderson and Stephen L. S. Smith

1. INTRODUCTION

IN 1994 the US and Chile officially began negotiations aimed at Chile's accession to the North American Free Trade Agreement (NAFTA). The US administration is also considering how to extend free trade to the whole of the Americas. The answer, according to administration officials, may be to bring in existing Latin American free trade areas. By negotiating with several nations at once the US seeks to encourage the necessary market reforms on would-be entrants to NAFTA (Greenhouse, 1994). For their part, members of these free trade areas hope to increase their leverage in accession negotiations by bargaining together. Thus NAFTA expansion beyond Chile will likely occur by adding the Andean Pact (Ecuador, Peru, Venezuela, Colombia and Bolivia) and MERCOSUR (Brazil, Argentina, Paraguay and Uruguay).[1]

From the new entrants' perspective, the long term success of FTA expansion depends on a variety of dynamic factors, such as the government's credible, public commitment to a liberal trade regime that an FTA connotes. The political battle in the US in approving any accession, however, is much more likely to focus on short-run adjustment costs. As with Mexico's accession, the debate is likely to include concerns for the initial employment and output losses in certain

---

MICHAEL A. ANDERSON is from Washington & Lee University, Lexington, VA. STEPHEN L.S. SMITH is from Gordon College, Wenham, MA. They thank seminar participants at the US International Trade Commission, the University of Western Ontario, and the Southern Economics Association meetings, and an anonymous referee, for helpful comments on earlier versions of this research. The authors also wish to acknowledge with gratitude the assistance on data matters of the staff of the Economics Research Service of the US Department of Agriculture, and of the Economics and Statistics Administration of the US Census Bureau.

[1] See Whalley (1996) and Rivera (1995) for more on accession strategies. Frenkel and Wei (1995) discuss whether regional trade agreements encourage or diminish the political impetus for larger multilateral liberalisation.

© Blackwell Publishers Ltd 1997, 108 Cowley Road, Oxford OX4 1JF, UK and 350 Main Street, Malden, MA 02148, USA.

import-sensitive industries. Ironically, therefore, the ability of the new entrants to exploit the long run benefits depends in part on whether US policymakers can anticipate the short-run concerns and negotiate a successful agreement. This paper's agenda is to shed light on the nature and extent of short-run changes in US imports resulting from NAFTA expansion.

We use three widely-cited partial equilibrium models, with disaggregated data, to investigate the near term effects of the potential new trade agreements with Latin America. Previous empirical work has estimated these models with inconsistent, often *ad hoc*, elasticity estimates. In contrast, we use disaggregated estimates of the relevant elasticities, and take advantage of the constraints on those elasticities implied by utility maximisation, as developed in Sawyer and Sprinkle (1989). As such, our estimates provide a more plausible range of outcomes than past partial equilibrium approaches.

Four scenarios are presented — a US-Chile FTA, a US-Andean Pact FTA, a US-MERCOSUR FTA, and, finally, an FTA encompassing US trade with all of the Latin American countries represented in the first three scenarios. Several important questions are addressed. First, what is the total effect of these proposed extensions of NAFTA on US merchandise imports? Second, how will the initial import surge be distributed across industries? Third, what are the welfare consequences for the US of these changes? Finally, what is the geographic distribution of affected industries? The answers will be particularly important for trade policymakers, providing a picture of the short-run trade and welfare consequences of NAFTA expansion for the US, and a profile of those import-competing industries likely to bear the brunt of any short-term adjustment.

We abstract from all considerations apart from the level of trade barriers themselves which influence the extent of trade between the US and Latin America. All long-run sources of trade expansion, such as economies of scale, changes in expectations about the permanence of policy changes, and changes in foreign direct investment, are set aside. We focus on the 'pure' trade effects on US merchandise imports of removing all tariffs with the relevant Latin American partners. While the long-run issues are critical for determining the eventual success of an expanded NAFTA, they are not relevant for our concern for the near-term consequences of expanded US preferential trade.

In each scenario we find that overall import changes are modest but the industry concentration of these changes is high. With a US-Chile FTA we estimate a mere 3.5 per cent increase in imports from Chile. Half of that increase, however, is in textiles and apparel. A US-Andean Pact FTA results in similar overall modest effects — the total increase in US imports is 2.2 per cent — and a similar concentration in textiles and apparel. Free trade with MERCOSUR and with all of Latin America is predicted to raise overall US imports from those regions by 9.1 and 4.7 per cent, respectively. The increase in imports from MERCOSUR will be strongly concentrated in leather footwear.

© Blackwell Publishers Ltd 1997

We predict that US welfare, measured by changes in trade creation and trade diversion, will increase under any proposed expansion of NAFTA. That welfare increase, however, is constrained by the small overall size of the trade expansion. Our predicted increase in imports from Chile under an FTA amount to 0.01 per cent of total US merchandise imports, and, for all of Latin America, 0.23 per cent of total US merchandise imports. In addition, the models used to estimate the welfare changes produce dramatically different estimates of the extent to which trade creation dominates trade diversion.

In the next section we describe our methodology and compare it with the major alternative, CGE modelling. Section 3 presents our main results, contrasts them with other findings, and considers how our projections may be affected by trade policy changes arising from the recently-completed Uruguay Round negotiations. The final section summarises our findings and their policy implications.

2. METHODOLOGY

a. Model Specification

We employ a variety of *ex ante* partial equilibrium methods to estimate the total change in imports by industry at the disaggregated SIC 3-digit level. The total change in imports is the sum of trade creation (TC) and trade diversion (TD).[2] Overall effects on US imports from each prospective free-trade partner (Chile, the Andean Pact, MERCOSUR, and the entire group of countries, respectively) are then found by aggregating the industry results.

Partial equilibrium approaches have been employed in this manner since the 1960s, and much is now known about their strengths and weaknesses.[3] The standard assumptions in these approaches are that markets are competitive, that income effects arising from price changes attributable to tariff cuts are negligible, and that within each industry domestic output, imports from the beneficiary of the tariff cut, and imports from the non-beneficiaries are imperfect substitutes for one another. The latter assumption is plausible on *a priori* grounds not only because of differentiated product considerations but because of factors associated with the ease or difficulty of transacting with particular countries or regions. Preferential tariff cuts then will serve to increase domestic demand for imports from the

---

[2] Viner (1950) distinguished between a tariff cut's welfare enhancing TC, the result of reducing the price to domestic users, and the tariff cut's welfare decreasing TD, the shifting of trade from non-beneficiary sources to beneficiaries. Imports from the beneficiary country increase by the sum of trade creation and trade diversion effects. The overall effect on the welfare of the country making the tariff cuts depends on the balance between these two effects.

[3] Balassa (1967), Verdoorn and Schwartz (1974), Kreinin (1973), Cline *et al.* (1978) and, most prominently, Baldwin and Murray (1977) all use partial equilibrium methods. Laird and Yeats (1990) discuss applications of the method. Theoretical issues are surveyed by Clague (1971 and 1972), Pomfret (1986b), McPhee (1987), and Sawyer and Sprinkle (1989).

© Blackwell Publishers Ltd 1997

beneficiary region at the expense of domestic output (by definition TC) and imports from non-beneficiaries (by definition TD).

Actually estimating TC and TD is a difficult empirical problem. Three elasticities are crucial in determining the results. The own-price elasticity of demand for the good receiving the preference influences the extent of the trade creation induced by the tariff cut; *ceteris paribus*, the greater this elasticity, the greater the trade creation. The elasticity of export supply in the beneficiary nation influences both trade creation and trade diversion. The greater this elasticity the greater will be the supply response to the tariff cut (i.e., TC) and to the shift in demand toward beneficiary goods away from non-beneficiaries (i.e., TD), *ceteris paribus*. Finally, the elasticity of substitution in consumption between preferred and non-preferred goods matters; the higher is this elasticity the higher is trade diversion, *ceteris paribus*.

In practice the preference receiving good's own-price elasticity is proxied by an aggregate measure of import demand elasticity.[4] It is also the norm to assume infinite supply elasticities. This not only simplifies the supply side of the model but vastly simplifies the cross-price effects: since only the beneficiaries' goods change price when the tariff is removed, no other cross-price effects muddy the demand side of the model. Assuming perfectly elastic supply somewhat overstates the amounts of TC and TD; rising marginal costs will constrain the actual increase in beneficiary imports. However, we show elsewhere that the *ratio* of TC to TD is unaffected by the fact that a beneficiary's supply elasticities are not infinite.[5] Accordingly we assume infinite supply elasticities across the board.

Details on our estimating equations are provided in the Technical Appendix. We use the standard equation for TC and three widely-cited equations for TD. The Verdoorn formulation of TD (Verdoorn, 1960; and Pomfret, 1986b) assumes that substitution takes place between goods produced by beneficiaries and non-beneficiaries, and between imports taken as a whole and domestic output. Alternatively, substitution can take place directly between domestic and beneficiary goods, and between domestic and non-beneficiary goods, as well as between beneficiary and non-beneficiary goods. This underlies Baldwin and Murray's (1977) equation for TD, widely adopted by subsequent researchers. The final estimating equation for TD comes from Laird and Yeats (1986). This has been used in Laird and Yeats (1990), Erzan and Yeats (1992), and in studies employing the World Bank-UNCTAD 'SMART' (Software for Market Analysis and Restrictions on Trade) model.

---

[4] Using a country's aggregate import elasticity of demand in place of the own-price elasticity of demand of imports from the beneficiary is in effect to assume that the substitutability between the domestic good and the beneficiary's good is the same as that between the domestic good and the non-beneficiary's good. See McPhee (1987) and Sawyer and Sprinkle (1989).

[5] This technical demonstration is available from the authors on request.

© Blackwell Publishers Ltd 1997

The Verdoorn and Laird-Yeats equations require explicit values for the elasticity of substitution, which is problematic. Sawyer and Sprinkle (1989) demonstrate that assumed values for the elasticity of substitution often violate the constraints implied by the first-order conditions for utility maximisation. We heed Sawyer and Sprinkle's cautions. Our results, using properly constrained estimates of this parameter, avoid the *ad hoc* nature of previous estimates of TD.

In sum: for each scenario (putting equation numbers from the Technical Appendix in parentheses) we estimate TC (1), Baldwin-Murray TD (2), Verdoorn TD (3), and Laird-Yeats TD (5). In the US-Chile FTA case Chile is the beneficiary, the US the preference granter, and the rest of the world the non-beneficiary. In the other scenarios the exact country mix of beneficiaries and non-beneficiaries changes from case to case, and so the estimates use appropriately aggregated country data, by industry, in each case.

b. *Model Suitability*

The basic strengths of the *ex ante* partial equilibrium approach are that it provides plausible, transparent estimates of trade diversion and trade creation at an appropriately disaggregated level, and that it is easy to assess the sensitivity of the estimates to alternative specifications of the key elasticities. The basic shortcoming of the approach is that it ignores the potential general equilibrium effects of the trade policy change, which in some cases may be substantial enough to affect the short-term, 'impact' effect on trade which we wish to estimate. Also, in the presence of large non-tariff barriers partial equilibrium approaches focusing on tariff removal only may underestimate the effects of complete liberalisation. Thus, before we turn to our estimates, we address the suitability of the partial equilibrium approach to US-Latin American trade liberalisation.

The major alternative methodology is the CGE approach, which has the inherent advantage of dealing with all of the general equilibrium repercussions of a policy shift.[6] However, CGE models have several shortcomings when applied to the short-run scenarios under evaluation in this paper. First, it is in the very nature of CGE modelling that the widely-used 'Armington assumption' of differentiated products by country generates large changes in terms of trade when trade liberalisation occurs. In effect, no country is allowed to be 'small,' which is unrealistic in general (see Brown, 1988)[7] and in these scenarios in particular. Second, as Summers (1991) documents, Latin America and the US are not

---

[6] Gravity models have also been used to evaluate the effects of trade liberalisation, with the preference margin included as an independent variable to predict trade flows, as in Sapir and Lundberg (1984), although this does not distinguish trade creation and trade diversion.

[7] The Armington assumption is not required in CGE models; output may be differentiated at the firm level, or there may be a mixture of firm and country differentiation, as in Brown *et al.* (1993). Francois and Shiells (1994) argue that the Armington assumption is usually the best choice.

© Blackwell Publishers Ltd 1997

'natural' free trade partners, lacking both geographic proximity and strong political/cultural ties. It seems plausible that the investment flows and other general equilibrium consequences of trade liberalisation will take place only after a substantial lag. Thus such CGE results may be properly thought of as long-term results. Finally, the high level of aggregation employed in CGE models may obscure important industry-level consequences.

In fact, several features of the liberalisation scenarios suggest that the partial equilibrium approach may be especially appropriate. The first feature is that the tariff barriers to be removed are relatively small and apply to small amounts of overall US trade. For instance, the share of Chilean goods in overall US merchandise imports — 0.3 per cent in 1991 — means that a US-Chile FTA is not likely to cause large general equilibrium effects in the US economy which would complicate the estimation of an FTA's consequences. In 1991 dutiable imports made up only 47.5 per cent of imports from Chile, by value, and faced an average tariff of only 3.1 per cent. Since the price changes engendered by the removal of this kind of low tariff structure will be small, this is an ideal context in which to use partial equilibrium analysis. This is also true of the other scenarios evaluated in this paper: average US tariff rates are a low 2.0 per cent for the Andean Pact and 3.9 per cent for MERCOSUR.

Furthermore, non-tariff barriers do not appear so high as to bias the results. In the Chilean case non-tariff barriers are virtually non-existent (see Erzan and Yeats, 1992; and Butelmann and Frohmann, 1992). Chile is not even covered under the Multifibre Arrangement (MFA). None of the Andean Pact countries is a signatory of the MFA (Peru's restrictions having expired in 1991), nor do Andean Pact countries face QRs in other sectors — although average MFN tariffs may be highest in important sectors such as non-rubber footwear and leather goods, there are neither quotas nor VRAs in these sectors. Brazil in 1991 was covered by the steel VRAs, but they were non-binding in this period (and, in any event, expired in 1992). The only major non-tariff barriers which may influence our results are found in MERCOSUR. Brazil, Argentina and Uruguay are restricted by the MFA.[8]

Taken together, neither partial equilibrium nor CGE approaches are free from problems in estimation. On balance, however, partial equilibrium models are well-adapted to measuring the short-run adjustment costs of the NAFTA expansion scenarios considered here.[9]

---

[8] A thorough survey of the extent and nature of US tariff and non-tariff barriers is presented in USITC (1993), which is the source for much of the information cited here.

[9] It is reasonable to ask at whose expense trade diversion will occur. Krueger (1995) argues that one problem with free trade areas as opposed to customs unions is that in the former the trade created upon the accession of a new member may come at the expense of the original beneficiaries. From the US point of view there are, in effect, two groups of non-beneficiaries when Chile or a country group joins NAFTA — Mexico and Canada on the one hand, and the rest of the MFN

© Blackwell Publishers Ltd 1997

3. DATA AND RESULTS

a. Data

The need to compare trade flows to US domestic shipments by SIC (in the Baldwin-Murray form of TD), for which 1991 is the most recent data available, and our desire to present a set of comparable estimates, requires working exclusively with 1991 data. Starting with US imports at the 10-digit HS level, we concord to the SIC and aggregate to the 3-digit level.[10] For Chile this yields 134 SIC 3-digit industries, and 143 and 150 SIC 3-digit categories for the Andean Pact and MERCOSUR, respectively.

We model the complete removal of tariffs in one step. In practice the FTA may remove tariffs more slowly over a multi-year period, and cuts may not proceed uniformly in all tariff lines. In this sense our estimates are upper bounds on the immediate trade effects of the FTA. The tariff rate for each SIC category is calculated as the ratio of total duties paid to the value of imports; in effect, we are using a weighted average tariff for the category, where the weights are the share of each 10-digit HS commodity in the 3-digit SIC category.[11]

We use plausible 'central' estimates of the relevant import elasticities of demand, supplemented with alternative estimates based on a range of possible elasticity values. Shiells, Stern and Deardorff (1986) present careful estimates for 122 3-digit SIC manufacturing industries. We use their point estimates when they have the correct sign and are statistically significant. For industries in which the Shiells et al. estimates were either of the incorrect sign or statistically insignificant we use the popular 'best guess' estimates of Stern et al. (1976). As these latter estimates were calculated for ISIC categories, we concord to the SIC to obtain a properly-weighted elasticity for each category in question. For agricultural goods we have taken the conservative step of assuming that import demand is unit elastic across the board.

---

world, on the other. Our estimates of TD make no distinction about where the diverted trade is coming from. This is reasonable because when tariffs on Chilean goods, say, are removed, the *relative* price change facing exporters in both groups of non-beneficiaries is the same. Barring large differences in the elasticity of supply between Mexico and the rest of the MFN world, TD will likely be distributed across all non-beneficiaries in proportion to their share of US imports; and the estimates of overall TD are likely to be accurate.

[10] Before aggregating to the 3-digit SIC level, a small portion of the lines of trade data was reclassified. For these lines, we used the principle differences concordance of the Commerce Department to match certain import SICs to domestic output SIC codes. The weights used were domestic shipment shares in those categories.

[11] In evaluating tariff removal at anything other than the tariff-line level an aggregation bias is introduced whenever the underlying tariffs on each commodity in the category are not equal. It can be shown that this bias is positive — i.e. the average tariff employed in the aggregated data will overstate the overall effects of the removal of the individual tariffs on the components of the category. This bias is small, however, when tariff changes are small. Again, it serves to give us an upper-bound estimate on the effects of tariff removal.

© Blackwell Publishers Ltd 1997

b. Results

For ease of reporting, the TC and TD calculations at the 3-digit SIC level have been aggregated to the 2-digit level. Table 1 presents the results from the four scenarios, the accession to NAFTA of, respectively, Chile, the Andean Pact, MERCOSUR, and a Western Hemisphere FTA including all countries in the first three groups. The largest six 2-digit SIC codes (in terms of trade creation) for each region are presented; these codes represent more than 75 per cent of the total trade effect for each scenario. (The Appendix contains disaggregated product information on the top ten 3-digit import expansion SIC codes for each region.)

Table 1 makes apparent that the Verdoorn TD — equation (3) — yields dramatically small estimates of TD compared to the Laird-Yeats equation (5), with the Baldwin-Murray estimate (2) lying somewhere in the middle. Differing assumptions between Verdoorn and Laird-Yeats dictate their wide divergence (discussed in the Technical Appendix). Accordingly, the fifth column of Table 1 reports the total trade effect (the sum of TC and TD) using the Baldwin-Murray estimate of TD — the median and, for reasons discussed in the Technical Appendix, our preferred estimate.

c. US-Chile FTA

Trade creation is $34.52 million, while trade diversion estimates are $17.0 million (Baldwin-Murray), $34.48 million (Laird-Yeats), and $0.4 million (Verdoorn). The median total trade effect of Chile's accession into NAFTA is $52 million. The overall impact of the FTA is therefore modest. Even assuming infinite supply elasticities, the total trade effect is only 3.5 per cent of current US imports from Chile.

The cases of textiles and apparel imports, however, bear special attention. Together they represent 57 per cent of all predicted trade effects. This is dramatic import growth — pre-FTA imports in these categories amounts to $46 million and the predicted increase is $29 million. Inside of overall small increases in trade is the possibility of such sectoral import surges.

d. US-Andean Pact FTA

Trade creation is $206 million, while trade diversion estimates are $105 million (Baldwin-Murray), $17 million (Verdoorn), and $205 million (Laird-Yeats). The total trade effect of the Andean Pact's accession to NAFTA is estimated at $312 million. This represents 2.2 per cent of the region's total merchandise trade with the US. As with Chile, complete liberalisation of trade with the Andean Pact results in only modest increases in trade.

Apparel (SIC 23) accounts for over half of the TC and, in the Baldwin-Murray and Laird-Yeats versions, over half of the TD. Combined with textiles the share

TABLE 1
Summary of Trade Creation and Trade Diversion by Region

(Top six SIC categories ranked by TC; figures in US$ millions; estimating equation from Technical Appendix in parentheses)

Region/SIC	*TC*	*TD*	*TD*	*TD*	*Total Trade Effect*	
	(1)	*(2)*	*(3)*	*(5)*	*(1)+(2)*	*% Share of Imports*
Chile						
23 Apparel	17.0	8.1	0.03	17.0	25.1	76.0
20 Food	4.1	0.3	0.09	4.1	4.5	1.7
22 Textiles	3.5	0.6	0.04	3.5	4.1	31.2
31 Leather	2.9	5.8	0.01	2.9	8.7	40.2
33 Prim. Met.	1.8	0.6	0.10	1.8	2.4	0.6
01 Crops	1.7	0.4	0.14	1.7	2.1	0.5
Other	3.5	1.1	0.01	3.5	4.6	1.5
Total	34.5	17.0	0.42	34.5	51.5	3.5
Andean Pact						
23 Apparel	108.2	56.2	1.6	107.4	164.4	57.7
13 Oil & Gas	25.1	12.3	3.4	25.1	37.4	0.6
29 Pet./Coal	20.1	10.9	6.2	20.1	21.2	0.6
01 Crops	14.1	9.7	5.0	13.7	23.8	2.9
22 Textiles	12.6	2.8	0.2	12.6	15.4	15.9
31 Leather	8.7	14.1	0.1	8.7	22.8	25.4
Other	17.5	9.2	0.4	17.5	26.6	0.8
Total	206.3	105.3	16.8	205.0	311.6	2.2
MERCOSUR						
31 Leather	143.7	242.6	19.8	140.8	386.3	30.3
33 Prim. Met.	62.1	10.7	4.5	61.7	72.7	6.4
20 Food	59.5	4.3	5.5	58.8	63.8	4.8
23 Apparel	45.8	25.6	0.4	45.7	71.4	36.0
35 Ind. Mach.	44.4	5.4	5.6	43.5	49.8	6.7
37 Trans. Eq.	21.3	7.7	0.1	21.3	29.0	4.3
Other	99.9	23.3	6.9	102.8	123.2	3.6
Total	476.7	319.5	42.7	474.6	796.3	9.1
Western Hemisphere						
23 Apparel	165.8	87.0	3.7	164.1	252.7	48.9
31 Leather	155.3	258.0	22.1	152.0	413.4	29.8
20 Food	71.8	4.3	17.9	70.6	76.1	3.0
33 Prim. Met.	68.2	11.8	5.9	67.7	80.0	3.7
35 Ind. Mach.	44.8	5.4	5.6	43.9	50.2	6.6
22 Textiles	37.3	6.8	2.5	36.5	44.2	14.0
Other	177.3	59.5	25.2	176.1	236.7	1.4
Total	720.4	433.0	83.0	710.9	1,153.4	4.7

Source: Authors' calculations.

© Blackwell Publishers Ltd 1997

of the total trade effect rises to 61 per cent. As with Chile, the importance of these sectors' trade effects is not driven by the fact that existing imports are large but because the import tariffs to be removed are relatively large and because the import elasticity of demand is also relatively large (−3.9 in several of the important 3-digit subcategories).

Venezuela's rich oil fields lead to the prominence of SICs 13 and 29, oil and refined oil products. Within the Andean Pact, however, Colombia would have the largest total trade effect — its net increase in exports of around $186 million amounts to 60 per cent of the Andean Pact's total increase. $125 million of this increase occurs in SIC 23, the apparel industry. Venezuela is second in the Andean Pact with a $64 million increase, or 21 per cent of the total trade effect.

e. US-MERCOSUR FTA

Trade creation is $477 million, while trade diversion estimates are $320 million (Baldwin-Murray), $43 million (Verdoorn), and $475 million (Laird-Yeats). The total trade effect of MERCOSUR's accession to NAFTA is estimated at $796 million, representing nine per cent of the region's total merchandise trade with the US. The presence of Brazil creates large predicted increases in exports for MERCOSUR, much larger than in Chile and the Andean Pact combined. And for the first time industrial products such as the output of primary metal industries (SIC 33) and industrial machinery (SIC 35) are among the leaders in TC and TD. From the US perspective, however, the increase of less than $800 million is still small relative to total imports of more than $500 billion.

Leather products dominate the total trade effect, here totalling $386 million or 48 per cent of the total. By far the largest 3-digit category inside SIC 31 is SIC 314, or footwear. This sector's prominence is due to a high level of initial imports from MERCOSUR (over $1 billion), a high initial tariff (10 per cent MFN), and an elastic import demand (−1.6). Brazil's exports of footwear are predicted to rise by $367 million, accounting for nearly all of the increase in MERCOSUR leather exports. Over all industry categories, Brazil accounts for $707 million, or 89 per cent, of the total trade effect.

The total trade effect in textiles (SIC 22) is $24.9 million, just low enough to keep it out of the 'top 6' categories listed in Table 1. It is worth noting, however, that the TC and TD calculations for MERCOSUR in textiles and apparel may be artificially low, given the effects of the MFA. An FTA which removes these quota restrictions along with the tariff barriers would almost certainly stimulate higher trade effects than we have estimated here (apart from consideration of the effects of the scheduled multilateral elimination of the MFA under the Uruguay Round agreement, taken up below).

© Blackwell Publishers Ltd 1997

f. US-Western Hemisphere FTA

Trade creation is \$720 million, while trade diversion estimates are \$433 million (Baldwin-Murray), \$83 million (Verdoorn), and \$711 million (Laird-Yeats). The total trade effect of Latin America's accession to NAFTA is estimated at \$1.15 billion. This represents 4.7 per cent of the region's total merchandise exports to the US, and 0.23 per cent of total US merchandise imports. Apparel, textiles, and leather products account for \$710 million, or 62 per cent of the total trade effect. Even here, however, the effects on import-competing US industries will be small, as it is only trade creation, not trade diversion, which comes at the expense of the domestic industry. In this scenario the 3-digit industry with the largest ratio of TC to domestic shipments is non-rubber footwear (SIC 314), at 3.9 per cent. The ratio of TC to domestic shipments among those 3-digit industries with the highest TC does not exceed 0.75 per cent in any other case.

g. Welfare Effects of FTA Expansion

Table 2 presents the ratio of TC to TD for the three scenarios using TD estimates from equations (2), (3) and (5). As noted earlier, while assuming an infinite elasticity of export supply biases the size of TC and TD upwards, the ratio of TC to TD is unaffected. Any proposed expansion of NAFTA to South America will increase the welfare of the United States. Every ratio in Table 2 is greater than 1; TC is larger than TD for every model and every scenario. What is less clear is the extent of any welfare improvement. The ratios vary by as much as 80 to 1 inside the same scenario, varying with the model used to estimate TD.

The Laird-Yeats estimates (column 4 of Table 2) suggest almost a 1:1 relationship between TC and TD; for every scenario TC is only slightly larger than TD. The Baldwin-Murray model (column 2) has TC between 1.5 and 2 times as large as TD. Finally, Verdoorn's model shows (column 3) the largest welfare gain with TC dominating TD by ratios between 8.7 and 81.2.

TABLE 2
Welfare Effects: Ratio of Total TC to Total TD, by Region
(Estimating equation for TD in parentheses)

Region	*TC/TD (2)*	*TC/TD (3)*	*TC/TD (5)*
Chile	2.0	81.2	1.001
MERCOSUR	1.5	11.2	1.004
Andean Pact	2.0	12.3	1.006
Western Hemisphere	1.7	8.7	1.013

© Blackwell Publishers Ltd 1997

h. Regional Effects

For simplicity we consider here the results of the US-Western Hemisphere free trade scenario only. Textiles, apparel, and leather footwear account for more than half of the total trade creation attributable to NAFTA expansion. Using the Department of Commerce's data on the distribution of sectoral value added by industry and state, we are able to suggest which states will be most affected by the short run increases in imports for these sectors. We assume that a state's share of the adjustment costs from increased imports will equal its share of total value added in that industry. So, for example, since North Carolina has 28 per cent of the total US value added in textiles, it will bear a similar per cent of the total adjustment from increased textile imports. North Carolina, Georgia, and South Carolina together will account for 61 per cent of total adjustment. In apparel the distribution of value added by state is less concentrated; the top five states, respectively, California, New York, Texas, North Carolina, and South Carolina will bear 33 per cent of the adjustment burden. In leather products including footwear, the five largest states, respectively, Missouri, Maine, New York, Wisconsin, and Massachusetts, again account for 33 per cent of the adjustment burden. However, these adjustment burdens will be small; as noted above, in all of the 3-digit sectors with large amounts of trade creation, only the leather footwear industry will experience trade creation in excess of 0.75 per cent of domestic shipments.

i. Comparison With Other Results

We are aware of two other major empirical studies of Western hemisphere free trade: Erzan and Yeats (1992), which uses a partial equilibrium approach, and the CGE modelling of Brown *et al.* (1993).[12] Erzan and Yeats, using 1986 data and the Laird-Yeats equation for TD, estimate that the total trade effect of free trade with Chile would raise US imports from Chile by 2.8 per cent. Free trade with all of Latin America (their equivalent of our Western Hemisphere FTA scenario) is predicted to raise US imports by 4.4 per cent.[13] These are slightly lower than our median estimates and are substantially lower than our estimates using the Laird-Yeats methodology (TC + TD(5) = 4.7 and 5.8 per cent, respectively). Further comparison is difficult because Erzan and Yeats neither specify their elasticity choices nor reveal disaggregated results.

Brown et al.'s CGE model is calibrated with 1990 data and estimates the effect of NAFTA accession, in sequence, of Chile, Argentina, Colombia, and Brazil.

---

[12] Butelman and Frohmann (1992) and Schiff and Ingco (1996) provide discursive treatments of NAFTA expansion.

[13] Erzan and Yeats (1992), Table 8, case with all US tariffs removed; we net out Mexico for this calculation.

© Blackwell Publishers Ltd 1997

Their results for Chile are comparable to our first scenario; their final estimation is close enough to our western hemisphere scenario to provide an instructive comparison. The model assumes 23 tradable product categories, rules out changes in FDI, and sets aside changes in NTBs to investigate changes in tariffs only. (By ignoring international capital flows, its predictions are in effect based solely on the long-term rationalisation of production dictated by goods price changes induced by the FTA.) Chile's accession to NAFTA is predicted to increase overall US imports (the total trade effect) by $427 million, a 32 per cent increase on 1990 imports from Chile.[14] In their final scenario, US imports from Argentina, Brazil, and Colombia rise by 8.5, 10.4 and 12.9 per cent, respectively. All in all these predictions are vastly higher than ours. In the US-Chile case, Brown et al.'s prediction is nine times larger than ours (3.5 per cent), while in the western hemisphere scenario their predictions range from two to three times as large as our 4.7 per cent. Their estimates are surprisingly large given the small change in tariffs under consideration in these scenarios. Given that Brown et al. rule out changes in FDI, the implied sectoral economies of scale necessary to generate these trade flow changes must be very large.

j. Implications of Uruguay Round for Projections

Finally, it is important to consider how the trade liberalisation begun in 1995 upon the completion of the Uruguay Round (UR) will affect the projections made in this paper. One central outcome of the UR was the agreement by the US to cut its average MFN tariff rate by 34 per cent, from a pre-UR trade-weighted average of 4.6 per cent to a post-UR average of 3.0 per cent. The implementation of these MFN cuts will lower the margin of preference accorded to any region which accedes to NAFTA. A rough sense of the consequences of this cut can be provided by assuming that the average 34 per cent tariff reduction is applied uniformly to the tariffs facing Chilean, Andean Pact, and MERCOSUR imports to the US. In this case all our estimates of trade creation and our central (Baldwin-Murray) estimate of trade diversion due to the removal of the (now lower) tariffs both fall by 34 per cent.[15] In the Western Hemisphere scenario, the predicted total trade effect would be an increase of 3.1 per cent in US imports rather than 4.7 per cent. Of course, in point of fact the UR agreement does not require the US to lower its tariffs uniformly. Textiles and apparel — sectors which we predict will see large growth under NAFTA expansion — are among the exceptions: the average apparel tariff will be 18 per cent even after the UR is fully implemented,

---

[14] Strictly speaking, not all of these new US imports come from Chile — some new US imports are induced from other areas as part of the general equilibrium ramifications of the trade regime change with Chile — although the preponderance will be from Chile.

[15] The other measures of TD fall less than proportionally in response to any given percentage change in the tariff cut; thus, in this case, they fall by something less than 34 per cent.

© Blackwell Publishers Ltd 1997

a cut of only 9.2 per cent, while the average textile tariff cut will be 24 per cent.[16] Thus in these important sectors the fall in the preference margin provided by NAFTA membership will not be nearly as great as 34 per cent. In this respect, again, our total estimated trade effect in these sectors in the Western Hemisphere scenario can be considered an upper bound on the actual effects attributable to the preference.

A second feature of the UR was that it tariffied agricultural quotas and sets out a six-year schedule for their reduction. However, with the exception of tobacco, this tariffication does not cover the major Chilean, MERCOSUR, or Andean Pact agricultural exports to the US. Before the UR these nations were essentially not bound by US agricultural QRs: the tariffied QRs are clustered in sugar and sugar products, dairy products, peanuts and peanut butter, and tobacco, rather than in the fruits and vegetables which have constituted the bulk of their agricultural exports to the US. Tariffication, then, neither affects this paper's projections nor suggests that substantial trade diversion will take place at Latin America's expense as the scheduled liberalisation proceeds.

Another interesting but somewhat speculative question posed by the UR is the that of the long-term consequences of dismantling the Multi-Fibre Arrangement (MFA). The MFA phaseout is significantly 'backloaded,' and hedged with elaborate safeguards, but nevertheless, promises to remove the quota restrictions by the year 2004 on US imports of textiles and apparel from potentially large exporters including Bangladesh, Thailand, and, assuming it joins the WTO, China. Will Latin American textile and apparel exports to the US be competitive, even with a NAFTA preference margin, in the face of these low-cost producers? Perhaps not. For the period up to 2004 our projections provide, we believe, an accurate picture of how US textile and apparel imports from Latin America will respond to accession to NAFTA. Beyond that, the fate of Latin American textile and apparel exports hinges on a number of imponderables such as the political and economic policies in the rest of the world which influence other exporters' ability to take advantage of the elimination of the MFA.

4. CONCLUSION

Our basic findings are as follows. First, under a range of plausible elasticity and theoretical specifications, the overall effects on US imports from NAFTA expansion to South America would be modest. Our median estimates are that the removal of tariffs on Chile increases imports from Chile by 3.5 per cent; with the Andean Pact the increase is 2.2 per cent, with MERCOSUR imports rise by 9.1 per cent, and a full US-Western Hemisphere FTA increases imports from the

---

[16] The details of the UR agreement cited in this paragraph come from Schott (1994).

© Blackwell Publishers Ltd 1997

region by 4.7 per cent. The change in total US imports would be almost imperceptible, lost in the noisy year-to-year volatility of trade. If anything, the implementation of the UR agreement over the next 10 years makes the predicted total trade effect of free trade even smaller, by reducing the preference margin inherent in removing (now smaller) MFN tariffs and by allowing other nations to compete in textile and apparel. Studies which show larger trade effects must be relying on predictions of changes in exchange rates, investment flows, large economies of scale, or other mechanisms to generate increased trade, as the effects of trade liberalisation *per se* are small.

Second, while overall increases in US imports from Latin America are small, certain sectors will see a significant increase in imports. Over a third of the total trade effect is found in just three industries — textiles, apparel, and leather footwear. This result reflects the unusually high level of protection, and the elastic US import demand, in these sectors. Because of the geographic concentration of US value added in these industries, we predict that most of the short run adjustment costs will also be concentrated in these same regions. In each scenario we find that at least 33 per cent of the predicted adjustment costs will occur in five or fewer states. But, crucially, these concentrated adjustment costs are likely to be small due to the small size of trade creation relative to existing domestic shipments in these industries. (In the Western Hemisphere FTA scenario, only in footwear does the predicted trade creation exceed 0.75 per cent of domestic shipments.) This may mean that, in contrast to the politically-contentious debate over Mexico's accession to NAFTA, free trade with Latin America may not spark intense opposition.

Third, the question of the likely growth in US exports to Chile, the Andean Pact, and MERCOSUR remains to be investigated and may, in fact, be significant in determining the strength of US political support for these FTAs. For instance, the removal of Chile's across-the-board 11 per cent tariff — a higher level of protection than facing Chilean exporters — may well stimulate more US exports to Chile than Chilean imports to the US are stimulated by US tariff removal. Since, in general, the tariffs faced by US exports to Latin America are greater than the tariffs imposed by the US on Latin American imports,[17] it may well be that under the other FTA scenarios considered in this paper US exports would rise more than US imports.

Finally, our results indicate a US welfare gain from NAFTA expansion. Yet the small size of the predicted trade changes means that any welfare change will be, at most, modest. This suggests that the question of Chile's, and the rest of Latin America's, accession to NAFTA needs to be evaluated in terms of its

---

[17] See Dean et al. (1994) for Chile and Brown et al. (1993) who report that average tariffs facing US exports to Argentina are 11.4 per cent, to Brazil 16.9 per cent, and to Colombia 10.5 per cent. Chile is scheduled to reduce its tariff to 7 per cent in 1997.

© Blackwell Publishers Ltd 1997

political significance and its significance as a signal about future economic policies in the NAFTA economies, rather than in terms of its immediate benefit to the US. If the agreement usefully signals continued US and Latin American commitment to open trade policies, if it supports the political and economic reforms in Chile, and if it serves to commit the rest of Latin America to recently begun reforms, then the agreement should be evaluated on those bases alone; its small immediate economic impact does not make the case one way or the other.

APPENDIX

Top 10 SIC 3-digit Import Expansion Categories, by Region

Ranked by TC
Total trade effect, $millions, measured as TC (1) plus Baldwin-Murray measure of TD (2).

Chile

(SIC)	($ million)	
233	14.5	Women's and misses' outerwear
232	10.2	Men's and boys' furnishings
203	3.4	Preserved fruits and vegetables
314	8.7	Footwear, except rubber
223	2.4	Broadwoven fabric mills, wool
333	2.2	Primary nonferrous metals
017	1.6	Fruits and tree nuts
214	1.2	Tobacco stemming and redrying
208	0.9	Beverages
222	0.9	Broadwoven fabric mills, manmade

MERCOSUR

(SIC)	($ million)	
314	378.1	Footwear, except rubber
331	70.7	Blast furnace and basic steel
203	53.7	Preserved fruits and vegetables
351	45.7	Engines and turbines
233	51.4	Women's and misses' outerwear
371	35.6	Motor vehicles and equipment
214	21.6	Tobacco stemming and redrying
286	13.8	Industrial organic chemicals
223	10.5	Broadwoven fabric mills, wool
201	5.6	Meat Products

Andean Pact

(SIC)	($ million)	
233	134.7	Women's and misses' outerwear
131	37.4	Crude petroleum and natural gas
291	20.6	Petroleum refining
018	23.1	Horticultural specialties
232	14.4	Men's and boys' furnishings
225	6.9	Knitting mills
314	14.1	Footwear, except rubber
391	11.0	Jewelry, silverware, and plated ware
236	5.3	Girls' and children's outerwear
286	4.1	Industrial organic chemicals

Western Hemisphere

(SIC)	($ million)	
314	396.5	Footwear, except rubber
233	200.0	Women's and misses' outerwear
331	75.1	Blast furnace and basic steel
203	64.3	Preserved fruits and vegetables
351	46.0	Engines and turbines
131	37.6	Crude petroleum and natural gas
291	24.4	Petroleum refining
371	28.7	Motor vehicles and equipment
214	22.9	Tobacco stemming and redrying
232	23.0	Men's and boys' furnishings

© Blackwell Publishers Ltd 1997

TECHNICAL APPENDIX

This Appendix presents and discusses the estimating equations. Subscripts 1, 2, and 3 refer to beneficiaries, non-beneficiaries, and the domestic (preference-granting) region, respectively. For any industry, define imports from beneficiaries as Q_1 and imports from non-beneficiaries as Q_2, both measured c.i.f. before the implementation of the trade preference scheme. For simplicity we suppress the industry subscript i. Define ε_m to be the elasticity of import demand. Call the initial tariff rate on imports from beneficiaries t. We use the standard equation for trade creation with infinite supply elasticities:[18]

$$\mathrm{TC} = Q_1 \varepsilon_m \frac{\Delta t}{1+t} \tag{1}$$

Note that in general trade creation rises as a proportion of initial imports the higher is the import elasticity of demand (in absolute terms) and the higher is the per cent reduction in the tariff. The absolute size of TC rises the higher are initial imports.

The Baldwin-Murray approach assumes that the elasticity of substitution between domestic and beneficiary output equals that between non-beneficiary and beneficiary output. This renders TD proportional to TC as follows:

$$\mathrm{TD} = \mathrm{TC}\left[\frac{Q_2}{Q_3}\right] \tag{2}$$

where Q_3 is domestic output.[19] The usefulness of this model is that it avoids the need for data on the elasticity of substitution between preferred and non-preferred imports. The Baldwin-Murray simplification has been criticised but in our view it provides, at the very least, a useful point of reference.[20]

In the Verdoorn formulation TD equals $Q_1\alpha_2(\varepsilon_s - \varepsilon_m)[\Delta t/1+t]$ where α_2 is defined as the share of non-beneficiary imports in total industry imports $(\alpha_1 + \alpha_2 = 1)$ and ε_s is the elasticity of substitution between beneficiary and non-beneficiary imports. Typically, empirical work employing this equation has assumed some constant value for ε_s.[21] Sawyer and Sprinkle (1989) argue that in practice, in the absence of taking explicit account of the restrictions imposed by

---

[18] See Baldwin and Murray (1977), Pomfret (1986b), and Sawyer and Sprinkle (1989).

[19] Baldwin and Murray (1977) and Sawyer and Sprinkle (1989) each present a derivation of equation (2).

[20] Pomfret (1986a) argues that (2) virtually guarantees that TC exceeds TD (which will be the case in any industry in which domestic output is larger than imports from beneficiaries). Baldwin and Murray (1986) defend their assumption as empirically plausible, and more recently Sawyer and Sprinkle (1989) argue that the Baldwin–Murray assumptions do not bias estimates of TD one way or the other against alternative specifications.

[21] Cline et al. (1978) use −2.5; values in other studies have ranged all the way up to −6.16, with many in the −2.5 to −3.0 range.

© Blackwell Publishers Ltd 1997

the first-order conditions of utility maximisation, researchers have tended to assume values for ε_s which are too high relative to ε_m. This has inflated estimates of TD. Leamer and Stern (1970) demonstrate that an upper bound for ε_s (in absolute value terms) is provided by ε_m/α_2. Thus Sawyer and Sprinkle make the substitution $\varepsilon_s = \varepsilon_m/\alpha_2$, in which case TD becomes:

$$\mathrm{TD} = \alpha_1 \mathrm{TC} \tag{3}$$

This 'restricted' version of the Verdoorn formulation will produce different estimates of TD than (2), depending upon the size of α_1 relative to Q_2/Q_3.

The final estimating equation for TD originates from Laird and Yeats (1986):

$$\mathrm{TD} = \frac{Q_1 Q_2 \varepsilon_s \left[\dfrac{\Delta t}{1+t}\right]}{Q_1 + Q_2 + Q_1 \varepsilon_s \left[\dfrac{\Delta t}{1+t}\right]}. \tag{4}$$

Note that in this formulation TD rises as ε_s rises, exactly as would be expected on the basis of theory.[22] But an explicit value for ε_s is needed to conduct the estimation; the SMART model uses -1.5 as the default value. It makes sense to apply the same substitution here as was done above in the Verdoorn formulation, yielding

$$\mathrm{TD} = \left(\frac{1}{1 + \varepsilon_m \dfrac{\alpha_1}{\alpha_2} \dfrac{\Delta t}{1+t}}\right) \mathrm{TC}. \tag{5}$$

Note that this renders TD proportional to TC, with the factor of proportionality depending upon the size of the tariff cut, ε_m, and α_1 (the share of beneficiary imports in total imports) relative to α_2 (non-beneficiaries' share). Since the factor of proportionality is always less than unity, TD will always be less than TC. In our results we estimate equation (5) and designate it the 'constrained' Laird-Yeats version.

As noted in the text, there is a wide divergence in the results based on (3) and (5). This is not surprising. Verdoorn's model multiplies TC by the share of beneficiary imports to obtain TD. For nearly every product, that share will be small for the country groups studied, resulting in the small values for TD. The Laird-Yeats construction, conversely, yields relatively large estimates of TD (close to but not exceeding TC) when tariff changes are small and the initial share of beneficiaries in total imports is small, as here.

---

[22] The derivation of (4) is not adequately explained in the literature; the best discussion is contained in Karsenty and Laird (1987).

© Blackwell Publishers Ltd 1997

Of the three models used to estimate trade diversion the Baldwin-Murray model has much in its favour. Though it has been criticised for making 'arbitrary' assumptions about the nature of substitution, when compared to suitably constrained versions of alternative TD measures it provides plausible results that are neither very small (Verdoorn) nor necessarily large (Laird-Yeats). Its construction with domestic output competing with both beneficiary and nonbeneficiary products, seems less restrictive than Verdoorn's assumption that competition occurs between beneficiary and non-beneficiary products, on the one hand, and between imports as a whole and domestic output, on the other. In the absence of a credible explication or defence of the Laird-Yeats approach in the literature, the Baldwin-Murray approach can be expected to retain its usefulness for the foreseeable future.

REFERENCES

Balassa, B. (1967), *Trade Creation and Trade Diversion in the European Common Market* (Yale University Economic Growth Center, Center Paper No. 103, New Haven, CT).

Baldwin, R. and T. Murray (1977), 'MFN Tariff Reductions and Developing Country Trade Benefits Under the GSP.' *Economic Journal*, **87**, pp. 30–46.

Baldwin, R. and T. Murray (1986), 'MFN Tariff Reductions and Developing Country Trade Benefits Under the GSP: A Reply,' *Economic Journal*, **96**, pp. 537–9.

Brown, D.K. (1988), 'Trade Preferences for Developing Countries: A Survey of Results,' *The Journal of Development Studies*, **24**, pp. 335–61.

Brown, D.K., A.V. Deardorff, D.L. Hummels and R.M. Stern (1993), 'As Assessment of Extending NAFTA to Other Major Trading Countries in South America' (Ann Arbor, MI: Institute of Public Policy Studies, University of Michigan, Processed).

Butelmann, A. and A. Frohmann (1992), 'US-Chile Free Trade,' in S. Saborio and contributors, *The Premise and the Promise: Free Trade in the Americas.* (New Brunswick, NJ: Transaction Publishers), pp. 179–94.

Clague, C.K. (1971), 'Tariff Preferences and Separable Utility', *American Economic Review*, **61**.

Clague, C.K. (1972), 'The Trade Effects of Tariff Preferences', *Southern Economics Journal*, **38**, 3.

Cline, W.R. et al. (1978), *Trade Negotiations in the Tokyo Round: A Quantitative Assessment*, (Washington, DC: The Brookings Institution).

Dean, J.M., S. Desai and J. Reidel (1994), *Trade Policy Reform in Developing Countries Since 1985. A Review of the Evidence*, World Bank Discussion Paper 267. (Washington, DC: The World Bank).

Erzan, R. and A. Yeats (1992), 'US-Latin American Free Trade Areas: Some Empirical Evidence', in S. Saborio and contributors, *The Premise and the Promise: Free Trade in the Americas* (New Brunswick, NJ: Transaction Publishers), 117–52.

Francois, J.F. and C.R. Shiells (1994), *Modeling Trade Policy. Applied General Equilibrium Assessments of North American Free Trade* (Cambridge: Cambridge University Press).

Frenkel, J. and S.-J. Wei (1995), 'Regionalism of World Trade and Currencies: Economics and Politics', Paper presented at the NBER Conference on Regionalisation of the Global Economy (October).

Greenhouse, S. (1994), 'US Plans Expanded Trade Zone', *New York Times* (4 February), D1.

Karsenty, G. and S. Laird (1987), 'The GSP, Policy Options and the New Round', *Weltwirtschaftliches Archiv*, **123**, 2.

Kreinin, M.E. (1973), 'The Static Effects of EEC Enlargement on Trade Flows', *Southern Economic Journal*, **39**, 559–68.

© Blackwell Publishers Ltd 1997

Krueger, A.O. (1995), 'Trade Creating and Trade-Busting Aspects of NAFTA', Paper presented at the American Economics Association meetings (January).

Laird, S. and A. Yeats (1986), *UNCTAD Trade Policy Simulation Model* (Geneva: UNCTAD).

Laird, S. and A. Yeats (1990), *Quantitative Methods for Trade Barrier Analysis* (New York, NY: New York University Press).

Leamer, E. and R. Stern (1970) *Quantitative International Economics* (Boston, Allyn and Bacon).

McPhee, C.R. (1987), 'The Consistency of Partial Equilibrium Estimates of Trade Creation and Trade Diversion', *Weltwirtschaftliches Archiv*, **123**, 1.

Pomfret, R. (1986a), 'MFN Tariff Reductions and Developing Country Trade Benefits Under the GSP: A Comment,' *Economic Journal*, **96** 534-6.

Pomfret, R. (1986b) 'The Effects of Trade Preferences for Developing Countries,' *Southern Economics Journal*, **53**, 18–26.

Rivera, S.A. (1995), 'After NAFTA: Western Hemisphere Trade Liberalization and Alternative Paths to Integration' (US International Trade Commission).

Sapir, A. and L. Lundberg (1984), 'The US Generalized System of Preferences and Its Impacts', in R.E. Baldwin and A.O. Krueger (eds.), *The Structure and Evolution of Recent US Trade Policy* (Chicago and London: University of Chicago Press).

Sawyer, W.C. and R.L. Sprinkle (1989), 'Alternative Empirical Estimates of Trade Creation and Trade Diversion: A Comparison of the Baldwin-Murray and Verdoorn Models,' *Weltwirtschaftliches Archiv*, **125**, 61-73.

Schiff, M. and M. Ingco (1996), 'Will Chile Benefit From Accession to NAFTA? Not Much' (The World Bank, February).

Schott, J.J. (1994), *The Uruguay Round: An Assessment* (Washington, DC: Institute for International Economics).

Shiells, C.R., R.M. Stern and A.V Deardorff (1986), 'Estimates of Elasticities of Substitution Between Imports and Home Goods for the United States,' *Weltwirtschaftliches Archiv*, **122**, 497-518.

Stern, R.M., et al. (1976), *Price Elasticities in International Trade* (London: Macmillan Press).

Summers, L. (1991), 'Regionalism and the World Trading System', in Federal Reserve Bank of Kansas City, *Policy Implications of Trade and Currency Zones.*

US International Trade Commission (ISITC) (1993), *The Economic Effects of Significant US Import Restraints*, Publication 2699 (November).

Verdoorn, P. J. (1960), 'The Intra-Bloc Trade of Benelux' in E.A.G. Robinson (ed.), *The Economic Consequence of the Size of Nations* (New York).

Verdoorn, P.J. and A.N.R. Schwartz (1972), 'Two Alternative Estimates of the Effects of EEC and EFTA on the Pattern of Trade,' *European Economic Review*, **3**, 291-335.

Viner, J. (1996), *The Customs Union Issue* (London: Carnegie Endowment for International Peace).

Whalley, J. (1996), 'Why Do Countries Seek Regional Trade Agreements?' NBER Working Paper 5552 (April).

DATA SOURCES

1. *US Imports of Merchandise International Harmonized System Commodity Classification (HTSUSA) by Country by Customs District* CD-IM-12, issued March 1992 by the US Department of Commerce, Bureau of the Census, Foreign Trade Division.
2. *1991 Annual Survey of Manufactures. Statistics for Industry Groups and Industries.* M91(AS)-1, US Department of Commerce, Bureau of the Census, December 1992.
3. *US Commodity Exports and Imports as Related to Output: 1991 and 1990* US Department of Commerce, Bureau of the Census, Economics and Statistics Administration, May 1994.
4. *Principal Difference Table* supplied by the US Department of Commerce, Bureau of the Census, Economics and Statistics Administration, August 1994.

© Blackwell Publishers Ltd 1997

703

49

WHARTON REPROGRAPHICS

NAFTA: Setting the Record Straight

.IBRARY
ELIVERY

Nora C. Lustig

1. INTRODUCTION

THE North American Free Trade Agreement (NAFTA) was sold on the basis that it would bring a substantial expansion of US exports to Mexico and boost Mexico's economic growth. Since its passage, however, not everything has gone as planned. In 1995 — scarcely a year after NAFTA came into effect — Mexico was in the midst of its worst economic crisis in decades. Meanwhile, rather than the anticipated expansion, US exports to Mexico went down.

Has, then, NAFTA failed? The short answer is 'no'. While many more years will be required before it is possible to make a meaningful assessment of the agreement, NAFTA's expected benefits are beginning to materialise. Trade among the three NAFTA countries has been expanding to record levels and the number of US-Mexican business partnerships is on the rise. Moreover, despite the 45 per cent real devaluation of the peso, the seven per cent drop in Mexican output and Mexico's 22 per cent fall in real wages during 1995,[1] US exports to Mexico were much less affected than Japanese or European exports largely thanks to NAFTA. Indeed, while exports to Mexico from the rest of the world, notably Japan and the European Union, fell by about 25 per cent, US exports contracted by less than two per cent.[2]

Despite these positive trends, many Americans in and out of the Congress are proclaiming that NAFTA was a mistake. To a large extent, the statements against NAFTA are based on erroneous beliefs or egregious distortions. NAFTA is blamed for US job losses or declining living standards of workers, particularly the unskilled. In reality, the impact of NAFTA on *gross* job displacement in the United States has been negligible. Furthermore, there is some evidence that NAFTA's *net* employment effect (the difference between jobs displaced and jobs created) has been positive.

NORA C. LUSTIG is Senior Fellow, Foreign Policy Studies Program, The Brookings Institution, 1775 Massachusetts Avenue, NW Washington, DC 20036, USA. This article is derived from: Nora Claudia Lustig, 'NAFTA: Setting the Record Straight', Policy Brief #20 (Washington: The Brookings Institution, June 1997).

[1] World Bank (1997).

[2] Banco de Mexico (1996).

© The Brookings Institution, 1997

World Economy 20(5): 605-614

Another argument used against NAFTA is the fact that following the Mexican peso crisis the United States is running a trade deficit with Mexico. However, unless it is the result of restrictions in market access, a deficit with any one country by no means is a cause of distress in the country that is running the deficit. The success of NAFTA has to be measured by the total amount of trade it creates regardless of which country is in deficit.

Finally, while some of NAFTA's critics argue that the agreement was a cause of the Mexican peso crisis, in reality the crisis was caused by factors unrelated to NAFTA. In fact, NAFTA is an important contributing factor to Mexico's economic recovery because of its impact on export performance and foreign direct investment flows.

These misperceptions probably reflect a deep-seated frustration with rising income disparities and declining real earnings of less-skilled workers in the United States for which trade with developing countries is frequently blamed. So far, the sentiment prevailing in Congress in relation to NAFTA and its extension has been strong enough to inhibit the administration from requesting Congressional approval for fast-track authority to negotiate the accession of other countries to NAFTA or any other free trade agreement. If current trends continue, it is unlikely that the United States will be able to remain a credible player — let along a leader — in the quest for hemispheric free trade.

To construct long-lasting support in the United States for NAFTA — and free trade more generally — action is needed on three fronts. First, the administration must undertake a systematic public education effort both with Congress and the general public about NAFTA and its actual effects. Second, further steps are required to make good on the promises imbedded in so-called labour and environmental side-agreements. NAFTA critics are right in one respect: there is no reason to forego NAFTA's contribution to the protection of workers' rights, particularly in Mexico, or improved environmental conditions, especially along the US-Mexican border. Finally, identifying ways to address the rising economic polarisation and declining living standards for the less-skilled in the United States deserves greater attention on the part of policy makers. Otherwise, many lawmakers are likely to continue focusing on the wrong instrument — that is, trade protection — to combat it.

2. NAFTA AND MEXICO'S ECONOMIC PERFORMANCE

Was NAFTA a cause of Mexico's 1995 financial crisis? Some observers have argued that it was. However, Mexico's large and rising trade deficit was not the result of the reduction in trade barriers implied by NAFTA but was caused by an appreciated peso (that is, the dollar was allowed to become too cheap). Furthermore, while the trade deficit contributed to the peso's devaluation, there

© The Brookings Institution, 1997

were other events that were equally, or more, important in explaining Mexico's financial debacle of 1995.

The Mexican financial crisis was the result of a combination of, on the one hand, rising US interest rates, political shocks, and market hysteria, and, on the other, erroneous policy responses on the part of the Mexican authorities. Three events in particular — the Zapatista guerrilla uprising in Chiapas, the assassination of the PRI presidential candidate, and the assassination of the PRI Secretary General — generated a climate of uncertainty among investors during 1994. In the aftermath of the PRI presidential candidate Luis Donaldo Colosio's assassination in March 1994, Mexico lost close to $10 billion — or about 40 per cent — of its international reserves.

While NAFTA cannot be blamed for the peso crisis, it is surely contributing to Mexico's recovery. And, economic growth in Mexico brings good things to the United States in the form of increased exports and reduced migration. The Mexican economic recovery got underway in the third quarter of 1995, and the 6.2 per cent fall in GDP exhibited that year was followed by a 5.1 per cent growth in output during 1996. Exports have been the engine of Mexico's recovery. Overall, they grew by 33 per cent in 1995, with non-oil exports rising by 47.5 per cent.[3] One indicator of NAFTA's role in this process is that the average rate of growth of Mexican exports to the United States was more than ten percentage points higher after 1994 than in the period 1991–1993. The sectors which showed the highest rates of growth were those where the regional integration was fairly advanced even before NAFTA: autos and auto parts, as well as textiles and apparels, food and beverages, and agriculture and cattle products.

NAFTA's impact is also observable in the behaviour of foreign direct investment, another important contributing factor to Mexico's economic recovery and future growth prospects. On average, between 1994 and 1996 foreign direct investment is almost double (close to $8 billion a year) the amount it was in the three years preceding NAFTA's implementation. Moreover, in 1995, when portfolio flows became highly negative, foreign direct investment was 75 per cent above the level in the three years preceding NAFTA. A comparison with the previous crisis may be telling. During 1983, foreign direct investment fell to a fifth of what it had been in the previous years. Clearly, the business opportunities brought by NAFTA have had a positive impact on foreign direct investment decisions, the peso crisis notwithstanding.

Unfortunately, the economic recovery has not yet reached the pockets of the average Mexican: between 1995 and 1996, real manufacturing wages fell cumulatively by over 25 per cent. But, if current economic trends continue — and current forecasts suggest they will — real wages should start growing, albeit

---

[3] INEGI Internet Site.

© The Brookings Institution, 1997

slowly, again. Still, it could take several years for Mexican wages to reach their pre-crisis levels but without NAFTA, it would have taken much longer.

3. NAFTA AND US TRADE

The relative size of the US economy *vis-à-vis* Mexico (about 25 to 1) will always limit the effects of NAFTA to very small orders of magnitude. Imports from Mexico as a share of total annual US imports, for example, have averaged about six per cent. Investment in Mexico is a tiny percentage of overall US investment. For example, in 1994, direct US investment in Mexico was $3 billion or about one-third of one per cent of US gross private domestic investment of more than $1 trillion.

Furthermore, assessing NAFTA's impact could turn into a futile exercise. The sharp devaluation of the Mexican peso and the crisis that followed clearly overshadow the effects of NAFTA's implementation. The devaluation and the recession caused a sharp decline in Mexico, US exports to Mexico, and Mexico's trade deficit with the United States of $2.4 billion in 1993 became a surplus of $15.4 billion in 1995. In addition, NAFTA is just one part of a continuing process of economic integration between Mexico and the United States, a process which accelerated with Mexico's trade liberalisation introduced in the mid-1980s. The benefits — and the costs — of economic integration will be spread over many years. Even in the longer run, the size of the US economy relative to Mexico will always limit the effects of trade and investment flows with that country.

Nevertheless, there are indications that NAFTA is making its imprint on both the magnitude and composition of bilateral trade flows. For example, during 1994, NAFTA's first year, trade between the NAFTA partners grew by 17 per cent over 1993, reaching a record of $350 billion, of which $100 billion are US–Mexico trade.[4] Taking a longer perspective, all of US–Mexico trade surged from $30 billion in 1986 (a year after Mexico began its trade liberalisation) to an estimated $140 billion in 1996.

In 1995, NAFTA worked to preserve and promote trade, despite Mexico's economic crisis. In contrast to the crisis in 1982, when Mexico re-introduced import permits for the majority of products, this time it kept its open-economy policies largely in place. Tariffs were increased to their GATT-bound levels in only a handful of sectors. But, because of NAFTA, the US — as well as other countries which had signed free trade accords with Mexico — were spared this tariff increase. Moreover, in 1995, while exports from the United States fell by about two per cent, exports from non-NAFTA countries — notably, Japan and the European Union — fell by 25 per cent. Part of the difference in export

---

[4] Weintraub (1997).

© The Brookings Institution, 1997

performance between the United States and other exporters to Mexico might be attributed to the depreciation of the dollar *vis-à-vis* other major currencies in 1995. However, most must be ascribed to the expanded market access for US exporters and the growing importance of intra-industry bilateral trade, both strengthened by NAFTA. NAFTA appears to have shielded US trade from the brunt of the Mexican crisis and the increase in some tariff lines that followed.

4. US EMPLOYMENT AND WAGES

Before NAFTA's approval, all opponents claimed that the agreement would lead to significant job losses in the United States. Cheap, labour-intensive Mexican imports, would outcompete US goods and plants would be attracted to Mexico by lower wages.

Three years of evidence reveals that those fears were unfounded. One indicator of the gross disemployment effects caused by NAFTA is the Trade Adjustment Assistance (NAFTA-TAA) certifications for all industries given by the Department of Labor. The *cumulative* figure between January 1994 and December 1996 equals close to 90,000 certifications. Sixty per cent of these are Mexico-related (as opposed to Canada or unspecified country). Forty per cent of this amount is due to plant relocation and the rest to employment displaced by imports. This number, however, is the total number of recipients of NAFTA-related assistance, which is not equivalent to the total number of people who may be unemployed as a result of NAFTA. USTR figures indicate that about 6,000 of the eligible recipients actually used the benefits, suggesting that the rest has probably been re-employed. Also, an accurate assessment of the net employment effect of NAFTA would need to look at job creation as well. A recent study which attempts to do this analytically found that the overall impact of NAFTA tariff liberalisation on US employment has been slightly positive.[5]

Moreover, job displacement may be causing real hardship for some individuals and families, but it is important to put the above figures in context. Since NAFTA, the United States has created in the order of 2.25 million jobs a year. Furthermore, according to the Bureau of Labor Statistics, total *gross* job displacements (subsequently re-employed in the majority of cases) during the first nine months of 1996 averaged 2.4 million workers per month. On a monthly basis, the NAFTA-TAA certified workers since the agreement came into effect equals 2,569 a month (82,223 divided by 32 months). NAFTA certified job displacements were only about one in one thousandth of the average monthly lay-offs during the first nine months of 1996.[6]

---

[5] Hinojosa et al. (1996).

[6] All these figures come from Hinojosa et al. (1996).

© The Brookings Institution, 1997

Given US wage flexibility, a more relevant measure of NAFTA-relate hardship for American workers may be its effect on wages, particularly for th less-skilled. While NAFTA's effect on wages has not been estimated, the impact of trade on US wages has been studied extensively. The majority of these studies concludes that international trade explains a small share of the observed rising wage inequality and downward trend in real wages of less-skilled US workers.[7,8] In general, it is technological change — with its bias in favour of high-skilled labour — that is the main factor behind these trends. It may be important to note that even if trade protection may bring small, short-term relief for unskilled workers, this will be at the expense of lower growth and living standards — including those of less-skilled workers — in the future. A more fruitful alternative would be to pursue policies designed to upgrade the education and skills of the working population.

5. LABOUR AND ENVIRONMENT

To respond to US congressional interest in incorporating labour and environmental issues into the negotiations, the Clinton administration negotiated the North American Agreement on Labour Cooperation and the North American Agreement on Environmental Cooperation, known as NAFTA 'side' agreements. Mexico and the United States alone established the Border Environment Cooperation Commission (BECC) and the North American Development Bank (NADBANK). The BECC:

> is expected to assist border states in both countries to design and finance environmental infrastructure projects certified by BECC and also to provide support for community adjustment and investment projects.[9]

As part of the so-called labour side agreement, a federal government level office in each country called the National Administrative Office (NAO) was created. One of its functions is to receive public complaints about non-enforcement of labour law.

The labour and environment NAFTA side-agreements were introduced to gain votes in the US Congress at the time of NAFTA's approval. However, organised labour, Democrats in Congress, and some environmentalists were not satisfied with the agreements because, in their words, they lack 'teeth': that is, no real provisions were made to use trade sanctions against violators of the side accords.

It is really fortunate that these agreements are 'toothless' in this respect. Otherwise, probably environmental and labour issues would have been frequently

---

[7] Burtless (1995).

[8] Lawrence (1996).

[9] GAO (1994, p. 11).

© The Brookings Institution, 1997

used as excuses to introduce protectionist barriers and obstruct NAFTA's implementation. Not even the European Union, so keen on harmonising social standards, uses trade sanctions in the face of labour or environmental law violations. In the European Union, complaints are brought before the judicial system by the interested parties in the country where the violation took place. The rulings have to be consistent with European Union norms, but all the remedial action is undertaken within the country so that the party who has been inflicted damage receives appropriate compensation.

Still, more could and should be done both on the labour and environmental fronts by the governments and NAFTA-created institutions. One repeated concern expressed by US labour leaders is that, given the characteristics of Mexico's political system and the corporatist role played by the official unions, Mexican workers still find it hard to form and belong to independent unions. This concern is well-founded. Even in today's more democratic Mexico, independent unions may face serious obstacles to be recognised. This appears to have been the case in several of the complaints brought before the US National Administrative Office. The recommendations of NAO, however, indicate an overcautious approach in dealing with this kind of labour rights violations.[10] Even if NAO's formal recommendations on these matters are not the equivalent of rulings and do not lead to sanctions (i.e., fines imposed on the country where the violations occurred) under the current terms of the side agreement, greater exposure of such violations could eventually lead to a reduction in their number. This would require a stronger stance on the part of the national NAO whenever such complaints are brought to its attention even if they pertain to another of the NAFTA members.

On the environmental front, the North American Development Bank (NADBANK) could be made an effective tool to foster environmental protection and clean-up, especially along the US–Mexican border. The Bank's capital will soon increase from $1.5 to $2.25 billion dollars. However, since its establishment, the Bank has approved only three loans for merely $2.98 million and issued guarantees for $1 million. One factor behind this low level of activity is that the established procedure — projects have to be certified by the BECC before they are analysed by NADBANK in terms of their financial viability — nay be flawed. Switching the order in which projects are analysed and approved nay expedite loans.

Another fundamental reason for this low level of lending is that the Bank's harter obliges it to operate in strictly commercial terms. Many of the potential orrowers (local governments, NGOs, etc.), particularly in Mexico, lack the dministrative experience and the ability to borrow funds or are simply not ligible because they would never be able to repay the loans required to finance a

---

Levinson (1996).

specific project. Environmental projects will need some form of subsidisation. NADBANK cannot lend at below-market rates, other sources, such a governmental non-refundable grants or private donations, will be needed to co-finance the projects with high social returns. Although this is already happening, the scale has been too small. As every other development bank, NADBANK may eventually need a window of credit on concessional terms for the poorest and most vulnerable communities. Furthermore, the bank should probably expand its portfolio of eligible projects and include housing, education, and poverty reduction projects more generally. This would not be in violation of its charter and such projects could have important environmental effects.

6. FINAL OBSERVATIONS

During the past three years, NAFTA has produced an increase in US–Mexican trade, business partnerships, specialisation in production processes, and direct investment flows into Mexico. At the same time, it has protected US exporters from the brunt of the Mexican crisis, especially in comparison to exporters from Japan and the European Union. While it is true that US exports to Mexico shrank during 1995, this fall was not generated by NAFTA but by the Mexican devaluation and the subsequent financial crisis. Furthermore, contrary to some beliefs, NAFTA was not a cause of the Mexican crisis and is now one of the pillars of that country's economic recovery. Finally, employment losses caused to American workers by NAFTA have been tiny despite the large rise in the US–Mexican wage gap caused by the peso crisis. There is even some evidence tha NAFTA has created more jobs than it destroyed.

Still, there remain a number of areas where progress is necessary. The United States and Mexico must contribute to the fulfillment and consolidation of the objectives set out in NAFTA. Several recent incidents underscore the need to deter the influence of powerful lobbies which may conspire against NAFTA's full implementation. An example of the latter in the United States is the decision to unilaterally overrule allowing the passage of Mexican trucks, which in effect constituted a violation of NAFTA. One way to avoid such situations would be to replace national trade remedy laws by an international competition code. This may insulate the trade disputes from political pressures of powerful vested interests. Since the US political climate is not ripe for giving up national trade remedy laws for internationally set codes, the responsibility to make good on NAFTA's commitments rests primarily with the US executive branch.

Additional efforts are necessary to make the institutions contemplated in the labour and environment NAFTA side-agreements more effective. In particular, the US National Administrative Office should be more forceful recommending action on the part of the Mexican government regarding cases where Mexican

© The Brookings Institution, 1997

workers' freedom of association is violated. In a country where the head of the Mexican equivalent of the AFL-CIO publicly threatens those who will not vote for the official party with retaliation, external pressure can play a significant role in improving the protection of workers' rights guaranteed by Mexican law, but not always in practice. The federal, state and local governments on both sides of the border should also step up their contribution to make more projects eligible for loans from the North American Development Bank. This will require making available non-refundable grants to co-finance projects which have a high social return. Given the proximity, helping Mexican communities to improve their living and environmental conditions can have large pay-offs for the United States.

In December 1994, at the Miami Summit of the Americas, the United States vowed to pursue the economic integration of the hemisphere and form a free trade area by the year 2005. Since then, however, little progress has been made. It is true that the hemispheric working groups created after the Summit are doing the groundwork for that objective to be achieved. But the administration has been slow in requesting fast-track negotiating authority in Congress to begin formal negotiations for Chile's accession to NAFTA. Chile has repeatedly manifested its interest to become a member of NAFTA and has completed free trade agreements with Mexico and Canada. However, its accession to NAFTA has become entangled in US domestic politics.

A laggard United States will not be devastating for economic integration in the hemisphere. However, because of its size, the US market is the biggest prize and, without that prize, the incentives to sustain open economy policies in some of the countries will diminish. Brazil may be a case in point. Brazil is the largest market in Latin America and, therefore, could be a significant lost opportunity for the United States should the process of liberalising trade and non-trade barriers between the two countries not take place. Furthermore, Central America and the Caribbean governments claim that being excluded from NAFTA has had negative effects on their economies already. These are countries whose prosperity and stability are also of vital interest to the United States because they are sources of migratory flows and a conduit for drug trafficking.

Now that the rest of the region is ready to move forward with economic ntegration, it is rather ironic that the United States after decades of promoting the principle of free trade appears hesitant. This image can only be removed if the Clinton administration moves forward with a request for fast-track negotiating authority and — together with its two partners — negotiates with Chile its accession into NAFTA. Free trade could well be like the proverbial bicycle. Either we continue to move forward or risk falling off.

Finally, the Mexican crisis underscores how macroeconomic events can stand n the way of reaping the benefits of greater economic integration implied by agreements such as NAFTA. The peso crisis points to the need to continue monitoring macroeconomic events in Mexico, and systematically analyse the

© The Brookings Institution, 1997

impact of US economic trends and policy decisions on the Mexican e While seeking full-scale macroeconomic coordination between the two c would be completely unrealistic at this point, a lot could be accompli continuing the close consultations between the authorities of the two c which developed in the context of the US financial rescue package. Whil has no right to interfere with or dictate Mexican macroeconomic policy versa), maintaining such consultations could help prevent another crisis the very least — give the US some warning that such a crisis might occ

REFERENCES

Banco de Mexico (1996), *The Mexican Economy, 1996: Economic and Financial Develo 1995, Policies for 1996* (Mexico City).

Burtless, G. (1995), 'International Trade and the Rise in Earnings Inequality,' *Journal of l Literature* (June), 33, 800–16.

Hinojosa, R. et al. (1996), *North American Integration Three Years After NAFTA: A Fram Tracking, Modeling and Internet Accessing the National and Regional Labor Marke* (University of California, Los Angeles, School of Public Policy and Social F December).

Instituto Nacional de Estadística Geografía e Informática (INEGI) (Internet Site), 'Si Cuentas Nacionales de Mexico (Mexico City).

Lawrence, R.Z. (1996), *Single World, Divided Nations*? (Brookings Institution Press an Development Centre).

Levinson, J. (1996), 'NAFTA's Labor Side Agreement: Lessons from the First Thre (Washington, DC: Institute for Policy Studies and International Labour Rights F November).

US General Accounting Office (1994), *North American Free Trade Agreement: Struc Status of Implementing Organizations* (Report No. GAO/GGD-95-10BR).

Weintraub, S. (1997), *Nafta at Three: A Progress Report* (Washington, DC: The Ce Strategic and International Studies).

World Bank (1997), *Mexico — Weekly Macroeconomic Indicators* (8 April).

© The Brookings Instituti

50

Has NAFTA Changed North American Trade?

David M. Gould
Senior Economist and Policy Advisor
Federal Reserve Bank of Dallas

Typically, the debate over NAFTA has focused on jobs. However, to really understand NAFTA's effects on employment or living standards, it is important to first answer the more fundamental question of what effect it has had on trade.

The North American Free Trade Agreement has been one of the most hotly debated trade accords in recent history. NAFTA's critics regard the expansion of free trade to a developing country like Mexico as a dangerous precedent. They envision U.S. jobs lost in a flood of goods from a country with an average wage one-fifth that of the United States. Others see NAFTA as a boon to U.S. employment and living standards through greater trade and investment opportunities.[1]

WHARTON REPROGRAPHICS

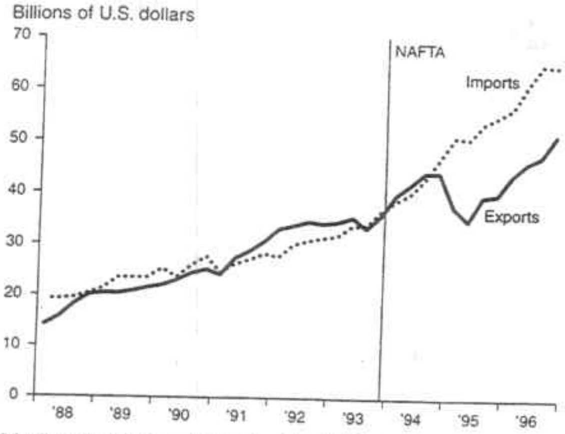
These opposing expectations for NAFTA have been largely matters of speculation and based on assessments of other trade accords. But now that NAFTA has been in operation for more than three years, the question is not what the trade accord is likely to do but what it has done.

Typically, the debate over NAFTA has focused on jobs. However, to really understand NAFTA's effects on employment or living standards, it is important to first answer the more fundamental question of what effect it has had on trade. Changes in trade patterns caused by a lowering of trade barriers are ultimately the mechanism by which jobs and living standards are influenced. This article examines how NAFTA, since its inception, has affected trade between the United States, Canada, and Mexico, holding constant other important factors that affect trade. Without controlling for these other factors, the effects of NAFTA are difficult to discern, which can lead to wrong or conflicting conclusions about the accord's effects on trade.

In the three years since NAFTA's implementation, both its supporters and opponents have used changes in the pattern of trade flows to justify their positions. Supporters have argued that during 1994, the year NAFTA took effect, U.S. trade with Mexico grew nearly 10 percent faster than the average of the previous five years (*Figure 1*). Opponents claim that any expansion in trade in NAFTA's first year was quickly reversed when expectations about its benefits fell to earth with the 1995 peso crisis. During 1995, U.S. imports from Mexico grew nearly 25 percent, but exports dropped 11 percent.

Since Mexico began to recover from its deep recession in late 1995, U.S. exports to Mexico have resumed rapid growth, but claiming the success or failure of NAFTA based on a superficial examination of the ups and downs of trade flows can be a mistake. The acccord's effects may be much more or less than a simple glance at these flows would suggest because NAFTA does not exist in an economic or policy vacuum.

Figure 1
U.S. Trade with Mexico



SOURCE: International Monetary Fund, Direction of Trade Statistics.

Worldwide economic changes that likely influenced bilateral trade within North America were already under way when NAFTA took effect. For example, U.S. real gross domestic product increased 3.5 percent in 1994, influencing the United States' supply and demand for imports and exports worldwide. As Figure 2 shows, U.S. trade with the world, excluding Mexico and Canada, grew faster in 1994 than in the previous six years. Likewise, Mexican real GDP increased 5.2 percent, and the real value of the peso was quite high in 1994, both factors that would have boosted U.S. exports to Mexico. As a result, it is unlikely that NAFTA and its lower trade barriers were the only influence on bilateral trade flows. To isolate the effects of NAFTA, one must account for the effects of changes in income, exchange rates, and trade with other countries; only then can NAFTA's impact on trade be ascertained.

In the first section of this article, I discuss previous analyses of NAFTA and compare them with my methodology. I next specify and estimate a model of the accord's effects on North American trade flows, and I assess how much trade has been influenced by NAFTA. I conclude by evaluating NAFTA's success.

DETERMINING NAFTA'S EFFECTS

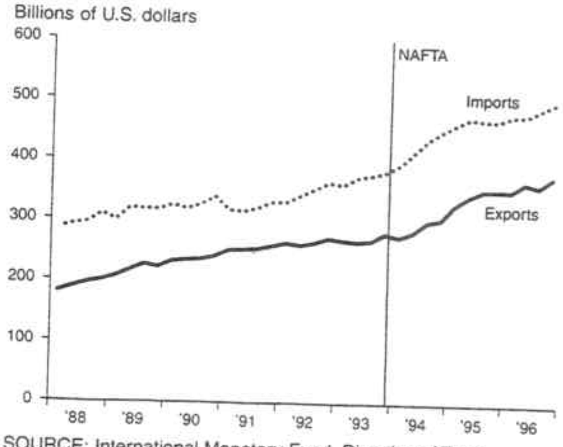
Because most studies of how NAFTA affects jobs, trade, and incomes in North America were completed just before, or shortly after, the treaty's implementation, the majority of them are forward-looking. This contrasts with the backward-looking approach of more recent studies, including this one.[2] However, any assessment of NAFTA's effects, past or future, must use an economic model to judge how trade interacts with the larger economy.

The techniques used to determine NAFTA's future impact fall into two broad categories. First is the computable general equilibrium (CGE) modeling technique, in which analysts model a simplified economy and simulate what would happen to it if tariffs and nontariff barriers fall according to the NAFTA schedule.[3] Typically, CGE models are static—that is, the structure of the economy cannot vary over time in response to changing trade patterns. More recently, however, a model by Kouparitsas (1997) allows for changes in capital investment and its reallocation across economic sectors and countries over time. Allowing capital flows to change over time in response to NAFTA generates a much larger benefit to freer trade.

Although these two types of studies make different assumptions about the economy's structure and the degree of competition in various sectors, both find an increase in income and trade under NAFTA. There is also substantial agreement about how NAFTA will affect Canada, Mexico, and the United States. In general, because Mexico's is the smallest economy, it reaps the largest benefit relative to its GDP (Kehoe and Kehoe 1994). The United States benefits modestly, and Canada does not have much benefit beyond that resulting from its 1989 free trade agreement with the United States.

Other studies use a partial equilibrium analysis to examine the effects of NAFTA. That is, they focus on particular sectors and assume

Figure 2
U.S. Trade with the World, Less Mexico and Canada



SOURCE: International Monetary Fund, Direction of Trade Statistics.

that the rest of the economy is unaffected. Determining NAFTA's trade effects in this context usually entails multiplying reduced-form price elasticities (a measure of how price changes translate into quantity changes) by the expected changes in tariff and tariff-equivalent trade barriers to determine how trade in various sectors is likely to change. In 1993, the U.S. International Trade Commission (USITC) completed a comprehensive study of NAFTA that used reduced-form price elasticities to determine how much NAFTA would affect trade in various U.S. economic sectors.[4] Hufbauer and Schott (1993) discuss NAFTA in a broad context and, based on their expectations of how trade would change under NAFTA, estimate how jobs in various sectors may be affected. They estimate changes in the number of jobs using Department of Commerce data on jobs supported directly and indirectly by exports to Mexico in 1990. They find that NAFTA and Mexico's economic reforms would increase the net number of U.S. jobs by about 170,000 in 1995 (Hufbauer and Schott 1993, 14).

In general, these partial equilibrium studies estimate that the United States will increase its exports of high-tech goods, grains and oilseed, and mechanical parts used in Mexican assembly plants. Increases are projected for U.S. imports of automobiles, apparel, glassware, household appliances, and certain horticultural products.

NAFTA SINCE ITS IMPLEMENTATION

Unlike the studies cited above, which have sought to predict how NAFTA will affect trade, this article's intent is to measure how NAFTA has already affected it. To assess the effects of NAFTA in its first three years, a model of bilateral trade flows in North America is empirically estimated with pre- and post-NAFTA data. The analysis is based on a widely used model of bilateral trade flows that includes incomes, prices, and exchange rates (see the box entitled "The Gravity Model of Bilateral Trade"). Once the fundamental determinants of trade flows are accounted for, any extraordinary flows that have occurred since NAFTA's inception are attributed to the free trade agreement.[5]

This type of analysis has at least two benefits. First, the analysis can estimate trade flows utilizing data since NAFTA's implementation. Second, it can capture most of the potentially important aspects of NAFTA, including tariff and nontariff barrier reductions, as well as changes in administrative rules, regulations, and expectations about the sustainability of free trade that cannot be easily quantified. Changing expectations for the sustainability of free trade under NAFTA are potentially among the most important aspects of the agreement. If NAFTA did not create a credible commitment to free trade, new investment would not flow into export industries to take advantage of reduced trade barriers. Without a credible free trade agreement, the benefits of the trade accord would be much lower. Expectations for a more stable and open trading environment affect trade by providing the incentive for firms to make long-term capital commitments. These expectations cannot be easily accounted for using the methodology cited previously because it is difficult to translate them into price changes.

Of course, there are potential pitfalls to this methodology as well. In particular, the model is unlikely to control for all the factors important to bilateral trade flows. Factors may be attributed to NAFTA that should not be. For example, when NAFTA took effect on January 1, 1994, political unrest was developing in Mexico, resulting in the armed Zapatista movement and two political assassinations later that year. Inasmuch as the uncertainty generated by this unrest reduced trade, it would reduce the estimated trade effects of NAFTA. Even though the events are independent of NAFTA, they would be indistinguishable in the context of the empirical model. If political uncertainty could be measured and included in the model, this would not be a problem.

However, even if important economic or political events are excluded from the model, they may not bias the estimated NAFTA effects if they and the accord's implementation were not simultaneous. Factors such as the peso crisis are unlikely to bias the analysis because exchange rates and incomes (the two factors most affected by the crisis) are included in the model. In other words, a majority of the peso crisis' effects on trade are likely to be taken into account.[6] Still, because NAFTA was implemented in a rather tumultuous period in Mexico that resulted in reduced trade, the estimated effects of NAFTA may be biased downward.

ESTIMATING THE EFFECTS OF NAFTA

To assess the effects of NAFTA since its implementation, the following benchmark model of Canadian, Mexican, and U.S. bilateral trade flows is estimated using quarterly data

The Gravity Model of Bilateral Trade

The empirical bilateral trade model in this article is based on the gravity equation, which derives its name from its resemblance to Newton's law of gravity. The model was originally formulated from ad hoc assumptions but had its intuitive appeal in describing trade flows as increasing as the economic "distance" between two countries shrinks. It describes bilateral trade flows between two countries as a function of their incomes, populations, the physical distance between them, and trade barriers. If countries with similar incomes have similar preferences for goods but produce different types of products, they are likely to trade more with each other than with other countries. Trade is also likely to increase the closer the countries are and the lower the trade barriers between them. The gravity model has been used to describe many different types of flows, such as immigration, shopping patterns, and car traffic, as well as interregional trade. It has been used extensively in international trade applications because it provides an empirically tractable framework.

The ad hoc assumptions behind the gravity equation have been replaced by microeconomic foundations. Anderson (1979), Helpman and Krugman (1985), Bergstrand (1985), and Bikker (1987) have developed variants of the gravity model based on utility and profit maximization. The empirical model this article uses is based on Bergstrand's theoretical foundation for the gravity equation, which is based on the assumption that producers maximize profits subject to a constant elasticity of transformation (CET) technology, and consumers maximize a constant elasticity of substitution (CES) utility function subject to a budget constraint (Bergstrand 1985). Assuming that individual bilateral trade flows are small relative to total trade, the equation for bilateral trade is

$$
\text{(B.1)}\quad
\begin{aligned}
PX_{ij} = {} & Y_i^{(\sigma-1)/(\gamma+\sigma)} Y_j^{(\gamma+1)/(\gamma+\sigma)} T_{ij}^{-\sigma(\gamma+1)/(\gamma+\sigma)} C_{ij}^{-\sigma(\gamma+1)/(\gamma+\sigma)} E_{ij}^{\sigma(\gamma+1)/(\gamma+\sigma)} \\
& \bullet \left(\sum_{k=1,k\neq i}^{N} P_{ik}^{1+\gamma} \right)^{-(\sigma-1)/(\gamma-\eta)/(1+\gamma)(\gamma+\sigma)} \\
& \bullet \left(\sum_{k=1,k\neq j}^{N} P_{kj}^{*1-\sigma} \right)^{(\gamma+1)(\sigma-\mu)/(1-\sigma)(\gamma+\sigma)} \\
& \bullet \left[\left(\sum_{k=1,k\neq i}^{N} P_{ik}^{1+\gamma} \right)^{(1+\eta)/(1+\gamma)} + P_{ii}^{1+\eta} \right]^{-(\sigma-1)/(\gamma+\sigma)} \\
& \bullet \left[\left(\sum_{k=1,k\neq j}^{N} P_{kj}^{*1-\sigma} \right)^{(1-\mu)/(1-\sigma)} + P_{jj}^{1-\mu} \right]^{-(\gamma+1)/(\gamma+\sigma)},
\end{aligned}
$$

where PX_{ij} is the value of aggregate trade flows from country *i* to country *j*,
Y_i is the aggregate income of country *i*,
Y_j is the aggregate income of country *j*,
P_{ik} is the price received for country *i*'s product in the k^{th} country,
$P^*_{kj} = P_{kj}T_{kj}/E_{kj}$ is the price paid for buying *k*'s product in the j^{th} market,
T_{ij} is 1 plus the ad valorem tariff rate on *i*'s product in the j^{th} market,
C_{ij} is a nontariff barrier on *i*'s product in the j^{th} market,
E_{ij} is the exchange rate between country *j*'s currency in terms of *i*'s currency,
γ is the constant elasticity of transformation in the supply between different export goods ($0 \leq \gamma \leq \infty$),
η is the CET between the supply of exports and domestically produced goods ($0 \leq \eta \leq \infty$),
σ is the constant elasticity of substitution between the demand for different imported goods ($0 \leq \sigma \leq \infty$), and
μ is the CES between the demand for imported and domestic goods ($0 \leq \mu \leq \infty$).

As shown in Equation B.1, the value of aggregate trade flows from country *i* to country *j* depends on nine terms. In the order of their appearance in the equation, they are (1) the income of the exporting country, (2) the income of the importing country, (3) tariffs, (4) nontariff barriers, (5) the exchange rate, (6) an export price index for exports to all countries to which the exporting country exports, (7) an import price index for imports from all countries from which the importing country imports, (8) an index of domestic prices for the exporting country, and (9) an index of domestic prices for the importing country.

These nine terms can be sorted into three categories: (1) income in the exporting and importing countries, which reflects the potential demand and supply; (2) wedges between the export and import price of the traded goods due to tariffs and nontariff barriers; and (3) price terms reflecting the substitutability between this traded good and the others.

Equation B.1 serves as the basis for the empirical model describing trade flows between the NAFTA countries. Changes in tariff and nontariff barriers are proxied by a binary variable for NAFTA. Because country-specific data for the price terms are not available, the empirical analysis uses proxies. To account for the exchange rate, domestic prices, and the terms of trade between the bilateral trade partners, GDP price deflators and the bilateral real exchange rate between the partners are used. To account for the terms of trade with other trading partners, a multilateral real exchange rate with the rest of the world is used. Economic events, such as the beginning of Mexico's trade liberalization in 1985 and Canada's free trade agreement with the United States in 1989, are proxied by binary variables.

from 1980 though 1996. The empirical equations are based on the gravity model, which is derived from standard microeconomic foundations (Bergstrand 1985). All variables are seasonally adjusted quarterly data and are expressed in log first-differences (growth rates):[7]

$$(1)\quad M_t^{ij} = \alpha_0 + \alpha_1 M_{t-q}^{ij} + \alpha_2 Y_{t-q}^{i} + \alpha_3 Y_{t-q}^{j} + \alpha_4 E_{t-q}^{ij} + \alpha_5 E_{t-q}^{iw} + \alpha_6 P_{t-q}^{i} + \alpha_7 P_{t-q}^{j} + \alpha_8 D_t + \alpha_9 NAFTA_t + \epsilon_t;$$

$$(2)\quad X_t^{ij} = \beta_0 + \beta_1 X_{t-p}^{ij} + \beta_2 Y_{t-p}^{i} + \beta_3 Y_{t-p}^{j} + \beta_4 E_{t-p}^{ij} + \beta_5 E_{t-p}^{iw} + \beta_6 P_{t-p}^{i} + \beta_7 P_{t-p}^{j} + \beta_8 D_t + \beta_9 NAFTA_t + \mu_t.$$

The variables are defined as follows: M^{ij} is country i's imports from country j; i and j are either the United States, Canada, or Mexico; t refers to the date; p and q are the number of periods a variable is lagged; X^{ij} is country i's exports to country j; Y^i is real GDP of country i; Y^j is real GDP of country j; P^i is country i's GDP price deflator; P^j is country j's price deflator; E^{ij} is the real exchange rate between countries i and j; E^{iw} is the real exchange rate between country i and the world (excluding country j); and D is a binary variable that represents changes in trade regimes not associated with NAFTA. For trade with Mexico, D equals 1 beginning in 1985, the period in which Mexico begins liberalizing trade. For trade between Canada and the United States, D equals 1 beginning in 1989, representing the period of the U.S.–Canada free trade agreement. *NAFTA* is a binary variable representing the period in which the accord was implemented. The variable *NAFTA* equals 1 beginning the last quarter of 1993.[8] α and β are estimated coefficients, and ϵ and μ are error terms.

The variables in Equations 1 and 2 can be sorted into four categories: (1) lagged trade (M_{t-p}^{ij}, X_{t-q}^{ij}), which reflects the adjustment process of trade to a new equilibrium; (2) income in the exporting and importing countries (Y^j, Y^i), which reflects the potential demand and supply for the traded goods; (3) price and real exchange rate terms (P^i, P^j, E^{ij}, E^{iw}), reflecting the substitutability of nontraded and traded goods in the NAFTA countries and the rest of the world; and (4) one-time trade liberalization variables, reflecting changes in trade regimes and NAFTA (*D*, *NAFTA*).

The size and statistical significance of the coefficient on the *NAFTA* variable tell us the degree to which NAFTA affects bilateral trade flows in North America.[9] It should be noted that NAFTA is not scheduled to be fully implemented until 2009. Tariff rates in many sectors are to be reduced over a fifteen-year period (see the box entitled "What Has NAFTA Done?"). Consequently, these results should be seen as a preliminary look into NAFTA's effects on aggregate trade.

The estimated equations are in the appendix.[10] Overall, the equations explain the growth of trade relatively well.[11] However, the effects of NAFTA on trade flows (in size and statistical significance) vary a great deal between countries. Figures 3 through 8 show what the estimation results imply for actual exports and imports between the United States, Canada, and Mexico. The shaded bands on both sides of trade estimated without NAFTA represent a 90 percent confidence interval derived from the statistical error of the estimate.[12]

The United States and Mexico

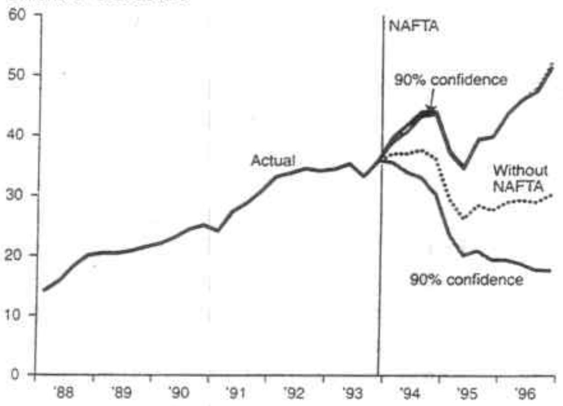
Figures 3 and 4 show NAFTA's estimated effects on bilateral trade flows between the United States and Mexico. As the dotted line in Figure 3 indicates, U.S. exports are estimated to have grown faster than they would have had there not been a trade agreement. On average, U.S. export growth is about 16.3 percentage points higher per year with NAFTA. While the increase in growth is not extraordinary, the cumulative effect is about $21.3 billion more in exports than what would have occurred without NAFTA. The statistical significance of this effect is high, as shown by the 90 percent confidence interval lines that exclude the observed data on U.S. exports to Mexico.

For U.S. imports, as shown in Figure 4, the boost from NAFTA is also relatively high. On average, import growth is about 16.2 percentage points higher per year with NAFTA. Since NAFTA became law, the cumulate impact amounts to about $20.5 billion in additional imports because of the agreement. However, NAFTA's statistical significance for U.S. imports from Mexico is only marginal. The 90 percent confidence interval lines show that we cannot exclude the possibility that trade without NAFTA would have been different from trade with NAFTA.

The United States and Canada

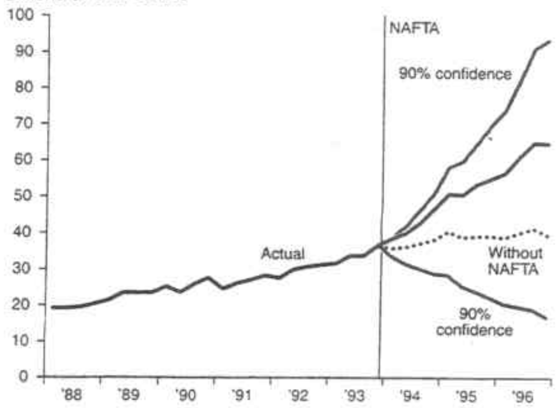
Figures 5 and 6 show the estimated effects of NAFTA on bilateral trade flows between the United States and Canada. As both figures show, trade between the two countries has not been affected much by NAFTA. This is not surprising, given that a free trade agreement with Canada was negotiated in 1989 and NAFTA did not alter

Figure 3
U.S. Exports to Mexico



SOURCES: International Monetary Fund, Direction of Trade Statistics; author's calculations.

Figure 4
U.S. Imports from Mexico



SOURCES: International Monetary Fund, Direction of Trade Statistics; author's calculations.

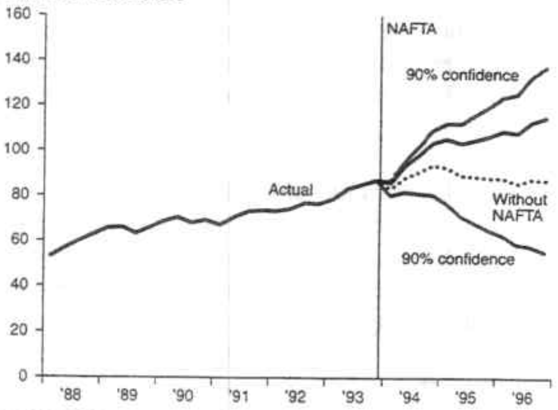
that accord significantly. However, even the 1989 agreement, as measured by the binary variable D^{ft} in the U.S.–Canada trade equations, does not seem to play a strong role in determining trade flows. Although trade may have been liberalized in some sectors, aggregate trade does not seem to be influenced much.[15] This may be because trade between the countries has been generally open for some time. Figures 5 and 6 show an 8.6 percent average annual increase in U.S. exports to Canada and a 3.9 percent increase in U.S. imports from Canada due to NAFTA. The NAFTA effect on both exports and imports is statistically insignificant.

Canada and Mexico

The effects of NAFTA on exports and imports between Canada and Mexico appear in Figures 7 and 8. As the figures show, the estimated effects are very imprecise, with a wide, 90 percent confidence band. One possible reason for the difficulty in measuring the effects of NAFTA in these equations is that trade between Canada and Mexico is a very small share of each country's total and is subject to much more unexplained volatility than is trade with the United States.[14] In these equations, the NAFTA trade effects are estimated to be negative, which raises the possibility that NAFTA may have diverted Canadian–Mexican trade toward the United States or other countries. But because the effects are so imprecise, the possibility that the effects are zero or even positive cannot be excluded.

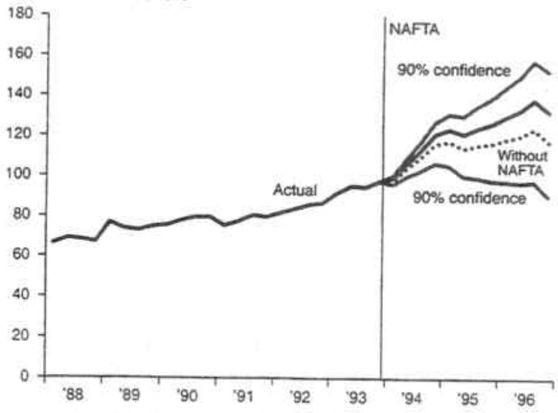
In summary, it is important to remember that there is a wide statistical margin of error for most of the estimated NAFTA trade effects, so they should be viewed in relative rather than abso-

Figure 5
U.S. Exports to Canada



SOURCES: International Monetary Fund, Direction of Trade Statistics; author's calculations.

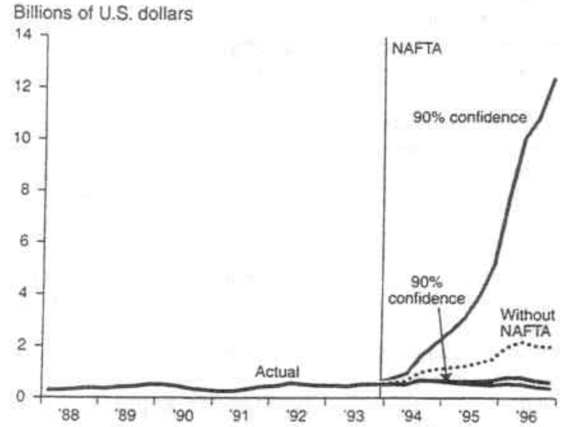
Figure 6
U.S. Imports from Canada



SOURCES: International Monetary Fund, Direction of Trade Statistics; author's calculations.

Figure 7

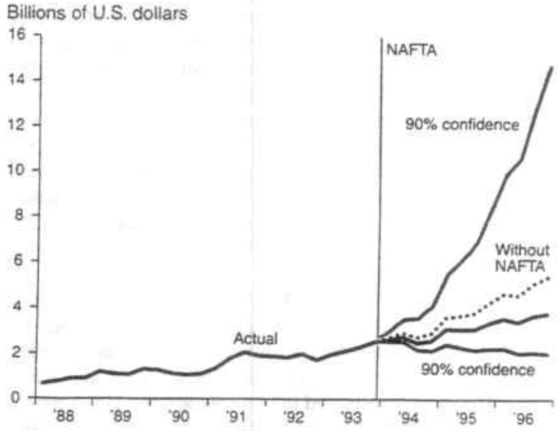
Canadian Exports to Mexico



SOURCES: International Monetary Fund, Direction of Trade Statistics; author's calculations.

Figure 8

Canadian Imports from Mexico



SOURCES: International Monetary Fund, Direction of Trade Statistics; author's calculations.

lute terms. While it is likely that NAFTA affected U.S.–Mexican trade, it is unlikely that it affected U.S.–Canadian or Canadian–Mexican trade.

TRADE DIVERSION VERSUS TRADE CREATION

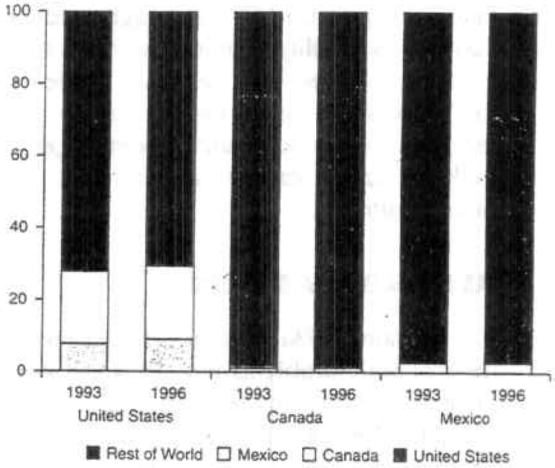
To judge NAFTA's effects on the economies of Canada, Mexico, and the United States, it is also important to consider North American trade flows in the context of trade with the rest of the world. In other words, did NAFTA create new trade opportunities within North America, or did it simply divert trade from countries outside NAFTA? If the increased trade caused by NAFTA was simply a shuffling of trade from other, more efficient trading partners, then NAFTA's benefit would shrink. Although a detailed examination of this issue is not within the scope of this article, a glance at how the distribution of trade flows has changed since NAFTA can tell us whether the accord may be associated with trade diversion.

Figure 9 shows how the distribution of trade flows between the NAFTA countries and the rest of the world changed from 1993, the year before NAFTA began, to 1996, three years after NAFTA started. As the figure shows, trade within North America has increased relative to trade with the rest of the world, but the increase is slight. The share of U.S. trade with Canada and Mexico increased from 27.8 to 29.4 percent between 1993 and 1996, with most of that increase attributed to greater U.S. trade with Mexico. Canadian trade with the rest of North America also increased, from a share of 77.3 to 80.4 percent. Mexico's trade share with North America changed very little, from 71 to 71.6 percent.[15]

The share of total trade between the NAFTA countries slightly increased, suggesting that if there was trade diversion, it was small. But to determine the extent of trade diversion in North America, it is also important to consider whether increased trade between NAFTA countries came at the expense of trade with the rest of the world. In other words, did NAFTA shift trade away from countries outside of NAFTA, or did NAFTA simply increase trade within North America at a faster rate than trade increased with the rest of the world? Figure 10 shows that it was the latter; that is, trade with countries outside North America also grew after NAFTA's implementation. The share of total trade between North American countries increased

Figure 9

Trade Shares of NAFTA Countries, Before and After the Treaty's Inception



SOURCES: International Monetary Fund, Direction of Trade Statistics; author's calculations.

What Has NAFTA Done?

On January 1,1994, NAFTA substantially reduced trade barriers across a market with more than 380 million people and a combined gross domestic product of roughly $7.6 trillion. Although trade barriers have already been lowered significantly, NAFTA will not be completely phased in until 2009. Most of the declines in tariffs have been on the Mexican side because Mexico started with higher tariffs than the United States or Canada. In 1992, Mexican tariffs on imports from the United States averaged about 10 percent when weighted by the value imported; at the same time, U.S. tariffs on imports from Mexico averaged about 4 percent. Because Canada and the United States negotiated a separate free trade agreement in 1989, NAFTA affected trade between the two countries very little.

NAFTA substantially reduces, but does not eliminate, nontariff trade barriers, such as import quotas, sanitary regulations, and licensing requirements. Canada and the United States traditionally have had few restrictions on capital flows, whereas Mexican laws prohibited private ownership in the petroleum industry and parts of the petrochemical industry, restricted foreign investment in the financial and insurance sectors, and institutionalized communal ownership of agricultural lands. The petroleum industry is still off-limits to foreign investment, although parts of the petrochemical industry are set to be privatized. Many laws against foreign investment in the financial and insurance sectors have been eliminated or substantially reduced. Although NAFTA set a schedule for liberalizing the banking sector, the 1995 peso crisis helped generate the political momentum to speed the opening of this sector by decreasing the restrictions on foreign ownership of existing banks.

NAFTA has not been without glitches, but the problems are probably fewer than what they would have been without the agreement. The opening of trucking between the United States and Mexico has been delayed, and tariffs have increased for some products. The United States levied additional tariffs on Mexican straw brooms, and in response Mexico levied tariffs on U.S. alcoholic beverages, flat glass, notebooks, and some types of wood furniture. There have also been disputes over agricultural products, such as avocados and tomatoes. Although these disputes are troublesome, their effect on overall trade has been small. Moreover, NAFTA may have limited a protectionist response to the 1995 peso crisis. Unlike Mexico's 1982 crisis, when the Mexican government raised tariffs dramatically in the hope of generating a trade surplus to boost foreign reserves, during the 1995 crisis no such political response occurred.

Here are some specifics by sector on how NAFTA has reduced trade barriers.[1]

Automobiles

NAFTA immediately decreased Mexican tariffs on automo s from 20 to 10 percent in 1994 and is set to drop them to zero by 2004. Tariffs on most auto parts will be eliminated by 1999. The agreement includes rules of origin specifying that to qualify for preferential tariff treatment, vehicles must have 62.5 percent North American content, which is an increase over the 50 percent provision in the U.S.–Canadian free trade agreement.

By 2004 NAFTA eliminates requirements that automakers supplying the Mexican market produce the cars in Mexico and buy Mexican parts. It has already eliminated mandatory export quotas on foreign-owned auto manufacturing facilities in Mexico, and by 1999 it eliminates the Mexican restriction on bus and truck imports.

Textiles and Apparel

NAFTA immediately eliminated trade barriers on more than 20 percent of Mexican–U.S. trade in textiles and apparel. Over six years it eliminates barriers on another 60 percent. The accord's rules of origin require that, to receive NAFTA tariff preferences, apparel be manufactured in North America from the yarn-spinning state forward.

Agriculture

NAFTA immediately reduced tariffs to zero for half of U.S. agricultural exports to Mexico. The other half of agricultural goods tariffs are to be eliminated by 2009. NAFTA immediately eliminated Mexico's licensing requirements for grains, dairy, and poultry.

Financial Services

NAFTA immediately reduced, and will eliminate by 2000, Mexico's restrictions on Canadian and U.S. ownership and provision of commercial banking, insurance, securities trading, and other financial services. Under NAFTA, Canadian and U.S. fina firms are allowed to establish wholly owned subsidiaries in Mex and to engage in the same range of activities as similar Mexican firms.

---

[1] Much of the following is described in Kehoe and Kehoe (1994).

---

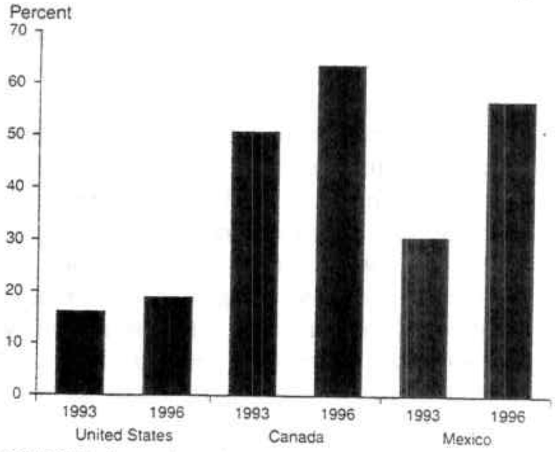
because trade within North America grew faster than did trade with countries outside of North America. Consequently, although trade diversion is a possibility, it is unlikely to be a large problem. Moreover, because trade under NAFTA was liberalized between countries with very different comparative advantages, it is unlikely that it caused a shift from optimal trading patterns.

HAS NAFTA BEEN A SUCCESS?

Certainly, NAFTA is not the solution to all the economic problems that ail North America, but it is not the disaster that critics claimed it would be. NAFTA is foremost a free trade agreement, and as such its benefits derive from a shift in resources to industries that reflect a nation's comparative advantage and away from indus-

Figure 10

Trade with the World as a Share of GDP (Excluding trade between NAFTA countries)



SOURCES: International Monetary Fund, Direction of Trade Statistics and International Financial Statistics.

tries that do not. It is important to understand that this shift implies that the benefits come from both increased imports and exports. Accordingly, the best way to judge a free trade agreement is by whether it increases imports and exports, and not by whether it increases exports and decreases imports. By this criterion, NAFTA has been a success for the United States and Mexico. As expected, NAFTA has meant little for the Canadian economy.

After accounting for the effects of economic variables important to bilateral trade flows—such as income, exchange rates, and prices—NAFTA is found to have a significant positive effect on trade flows between the United States and Mexico. NAFTA is not found to have a significant impact on trade between the United States and Canada or Canada and Mexico. These findings are not surprising, given that the United States negotiated a free trade agreement with Canada five years before the implementation of NAFTA and that most of the trade liberalized under NAFTA is between the United States and Mexico.

Although this empirical analysis controls for economic shocks that would affect trade through changes in incomes and exchange rates, such as the 1995 peso crisis, it cannot control for all external shocks, nor can it capture all aspects of NAFTA's influence on trade. Perhaps the largest omission from the analysis is trade barriers that were not erected because of the free trade agreement but would have been without it. This issue was particularly relevant during the 1995 peso crisis. Unlike previous periods of economic turmoil in Mexico, trade was relatively unimpeded during the peso crisis. NAFTA, by enhancing the economic ties between the North American countries, may have limited a protectionist response to the peso crisis and helped facilitate a return of foreign investment and economic growth to Mexico.

Has NAFTA destroyed U.S. jobs? Clearly, NAFTA has neither spelled the death of the U.S. workforce, nor has it generated a dramatic increase in the number of U.S. jobs. What dominates the employment picture in any year are movements in a country's own business cycle, not trade. U.S. income grew fairly smoothly between 1994 and 1996; as a result, U.S. employment grew by 3.6 million. In contrast, Mexico experienced a currency crisis and deep recession in 1995; its employment fell but is now recovering with the economy. Ultimately, freer trade does not determine the number of jobs available in a country, but it does determine the types of jobs available. In the three years since NAFTA's implementation, there has been a clear trend toward increased trade in North America and higher productivity in the United States. How much of that greater productivity is due to NAFTA is unknown. As time passes, and more economic data become available, cyclical factors and economic shocks will fall to the background and a clearer picture of NAFTA's effects on the economy will emerge.

NOTES

I thank Baoyuan Wang for excellent research assistance, and Evan Koenig, Bill Gruben, and Lori Taylor for their comments and suggestions. Any remaining errors are my own.

[1] The early controversy can perhaps be best summarized by quotes from Ross Perot and President Bill Clinton during the NAFTA debate in 1993: "NAFTA will pit American and Mexican workers in a race to the bottom. In this race, millions of Americans will lose their jobs" (Perot 1993, *i*); "I believe the Nafta will create 200,000 American jobs in the first two years of its effect" (Clinton 1993).

[2] See, for example, Gould (1996), Weintraub (1997), and USITC (1997).

[3] For an excellent survey of general equilibrium models applied to NAFTA, see Kehoe and Kehoe (1994).

[4] In creating the reduced-form price elasticities, the USITC study assumes that foreign and domestic goods are imperfect substitutes for each other. In other words, the goods have separate markets in which equilibrium prices and quantities are established. See USITC (1993) for a description of this methodology.

[5] The methodology used here is an extension of the work done by Gould (1996). Other recent studies have used a similar methodology to assess the trade and sectoral effects of NAFTA. See USITC (1997).

[6] See Gould (1996) for an assessment of how the peso crisis affected trade independent of NAFTA. See Neely (1996) for a discussion of why NAFTA did not cause Mexico's peso crisis.

[7] Log first-differences, as opposed to a simple log-linear relationship, were used because tests on the dependent and many of the independent variables could not reject the hypothesis of nonstationarity. Consequently, the equations estimate the growth of exports and imports.

[8] Because trade growth equations are estimated, the effects of NAFTA are assumed to influence the growth of trade. However, according to traditional long-run models of trade, lower tariffs only influence the level, not the growth, of trade. Because trade is unlikely to jump to a new, higher level instantaneously, the growth of trade is likely to be affected in the transition to a

new, higher level. This is especially true for the short period that NAFTA has been observed and because NAFTA is being phased in over fifteen years.

[9] NAFTA may also affect bilateral trade flows indirectly through income and prices. Although these indirect effects are likely to be important over the long run, over the short run these effects are probably small. Because of this, these secondary effects are ignored in the estimation.

[10] The equations were estimated with ordinary least squares and the errors terms checked to see if they follow a white-noise pattern. The lag structure of the equations was determined according to the Akaike information criterion. To determine how trade would have grown without NAFTA, the estimated *NAFTA* effect was excluded from the estimated exports and imports equations, and trade flows were calculated with the actual data for the independent variables. To provide the best estimates, the error term (which reflects the degree to which the equation does not match the data) was included in the calculation. Data sources are given in the appendix.

[11] The adjusted R^2 on the equations varies from 0.67 in the U.S.–Mexico export equation to 0.18 in the Canada–Mexico import equation. Most of the equations have an R^2 between 0.30 to 0.40, which is not uncommon for similar growth equations. The adjusted R^2 measures the proportion of the variation in the left-side dependent variable that is explained by the right-side dependent variables, adjusting for the number of variables in the equation.

[12] The confidence interval shows the degree of certainty we can have in the estimated effects. If the confidence interval around the estimated effects of trade without NAFTA excludes the actual observed trade under NAFTA, we can say with 90 percent certainty that trade with NAFTA is different from trade without it. If the 90 percent confidence interval includes the observed trade under NAFTA, we can say that there is less than a 90 percent certainty that trade is different with NAFTA than without it.

[13] An inherent problem in studying aggregate exports and imports is that the analysis cannot explain changes in sector-specific trade flows caused by NAFTA. For example, imports in one industry may expand, while imports in another industry may contract. In aggregate, however, imports overall would appear to remain stable. An attempt was made to study sector-specific trade data, but because equivalent sector-specific price information across countries does not exist, the empirical results were poor.

[14] This is indicated by the relatively low adjusted R^2 of the Canada–Mexico trade equations.

[15] A bilateral trade intensity index, defined as the share of country j's trade in country i's world trade relative to the share of country i's world trade in total world trade $[I_{ij} = (T_{ij}/T_{iw})/(T_{jw}/T_w)]$, also shows a slight increase among NAFTA partners since 1993. See Yeats (1997) for a discussion of this index applied to Mercosur's trade.

REFERENCES

Anderson, James (1979), "A Theoretical Foundation for the Gravity Equation," *American Economic Review* 69 (March): 106.

Bergstrand, Jeffrey (1985), "The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence," *Review of Economics and Statistics* 67 (August): 474.

Bikker, Jacob A. (1987), "An International Trade Flow Model with Substitution: An Extension of the Gravity Model," *Kyklos* 40 (3): 315–37.

Clinton, Bill (1993), quoted in Bob Davis, "Two Years Later, the Promises Used to Sell Nafta Haven't Come True, But Its Foes Were Wrong, Too." *Wall Street Journal*, October 26, 1995, A24.

Gould, David M. (1996), "Distinguishing NAFTA from the Peso Crisis," Federal Reserve Bank of Dallas *Southwest Economy*, September/October, 6–10.

Helpman, Elhanan, and Paul Krugman (1985), *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy* (Cambridge, Mass.: MIT Press).

Hufbauer, Gary C., and Jeffrey J. Schott (1993), *NAFTA: An Assessment* (Washington, D.C.: Institute for International Economics).

Kehoe, Patrick J., and Timothy J. Kehoe (1994), "Capturing NAFTA's Impact with Applied General Equilibrium Models," Federal Reserve Bank of Minneapolis *Quarterly Review*, Spring, 17–34.

Kouparitsas, Michael A. (1997), "A Dynamic Macroeconomic Analysis of NAFTA," *Economic Perspectives* 21 (January/February): 14–35.

Neely, Christopher J. (1996),"The Giant Sucking Sound: Did NAFTA Devour the Mexican Peso?" Federal Reserve Bank of St. Louis *Review*, July/August, 33–47.

Perot, Ross (1993), *Save Your Job, Save Our Country: Why NAFTA Must Be Stopped—Now!* (New York: Hyperion).

U.S. International Trade Commission (USITC 1997), "The Impact of the North American Free Trade Agreement on the U.S. Economy and Industries: A Three-Year Review," USITC Investigation no. 332–81, June.

U.S. International Trade Commission (USITC 1993), *Potential Impact on the U.S. Economy and Selected Industries of the North American Free-Trade Agreement*, USITC Publication 2596 (Washington, D.C.: USITC, January).

Weintraub, Sidney (1997), *NAFTA at Three: A Progress Report* (Washington, D.C.: Center for Strategic and International Studies).

Yeats, Alexander (1997), "Does Mercosur's Trade Performance Raise Concerns About the Effects of Regional Trade Arrangements?" World Bank Policy Research Working Paper no. 1729 (Washington, D.C., February).

Appendix
Regression Results

Dependent variable	U.S.(i)–Mexico(j)		U.S.(i)–Canada(j)		Canada(i)–Mexico(j)	
	Exports	Imports	Exports	Imports	Exports	Imports
Constant term	.027	–.061	–.025	–.029	–.229	.169
	(.794)	(.713)	(.688)	(.078)	(.224)	(.042)
Lagged dependent variable	–.679	–.901	–.441	–.209	–.299	–.323
	(.074)	(.050)	(.659)	(.098)	(.341)	(.012)
Y^i	3.623	.289	3.953	2.793	9.474	2.560
	(.031)	(.209)	(.334)	(.135)	(.366)	(.543)
Y^j	.082	–3.046	–1.095	.104	3.677	.335
	(.050)	(.539)	(.394)	(.425)	(.255)	(.841)
E^{ij}	–2.147	–.610	–.171	.178	–.863	–.435
	(.000)	(.317)	(.536)	(.427)	(.053)	(.181)
E^{iw}	–2.971	–.061	–2.062	–.418	6.748	5.742
	(.047)	(.316)	(.430)	(.665)	(.497)	(.007)
P^i	.963	13.716	3.438	4.208	–17.100	–5.847
	(.616)	(.292)	(.106)	(.009)	(.384)	(.149)
P^j	–.533	–.191	.477	–.877	–1.185	–.890
	(.704)	(.608)	(.724)	(.578)	(.368)	(.099)
D^{ij}	–.029	.026	–.005	.008	–.008	–.003
	(.522)	(.722)	(.825)	(.550)	(.951)	(.994)
NAFTA	.073	.072	.031	.018	–.111	–.038
	(.015)	(.119)	(.191)	(.374)	(.240)	(.453)
Adjusted R^2	.67	.35	.36	.31	.27	.18
Equation F statistic (significance level)	.000	.030	.080	.002	.120	.050
LM test for autocorrelation (significance level)	.31	.53	.20	.25	.01	.24
Lag structure	3	5	5	1	4	1
Degrees of freedom	30	14	18	50	22	46

NOTE: Coefficients are the sum of the lagged terms. Significance level of F statistics (the null hypothesis that all lagged coefficients are equal to zero) are in parentheses.

Data Sources

Variable	Definition	Source
M	Seasonally adjusted value of merchandise imports, in millions of U.S. dollars	International Monetary Fund, Direction of Trade Statistics
X	Seasonally adjusted value of merchandise exports, in millions of U.S. dollars	International Monetary Fund, Direction of Trade Statistics
P	Seasonally adjusted GDP price deflator	International Monetary Fund, International Financial Statistics
Y	Seasonally adjusted real GDP	International Monetary Fund, International Financial Statistics
E	Seasonally adjusted real exchange rate	Trade-Weighted Value of the Dollar, Federal Reserve Bank of Dallas, and author's calculations
E^w	Seasonally adjusted real exchange rate with rest of the world	Trade-Weighted Value of the Dollar, Federal Reserve Bank of Dallas, and author's calculations

LIPPINCOTT LIBRARY
DOCUMENT DELIVERY SERVICE

51

WHARTON REPROGRAPHICS

Riding the Russian Roller Coaster: U.S. Firms' Recent Experience and Future Plans in the Former USSR

Daniel J. McCarthy Sheila M. Puffer Peter J. Simmonds

Dramatic changes in the former Soviet Union have turned business conditions there into a roller coaster ride. The unstable situation has created a great deal of uncertainty for firms that look to this part of the world as a market for goods and services. Yet, the size of the market and the potential for international expansion remain tempting. To assess how the changing environment in the former USSR is affecting business decisions in U.S. companies, we conducted a survey in 1992 of forty-two companies (based primarily in New England) and also interviewed sixteen of the respondents in depth. This article reports on these companies' current activities, their plans for the future, and assessments of risk in doing business in the former USSR.

A Brief History of American Business Activity in the Former USSR

The Communist Regime—Throughout most of the communist regime, foreign investment was prohibited by the communist authorities who had an iron grip on the Soviet Union from the 1920s through the 1980s. Although considerable foreign capital, technology, and know-how were used to build and expand the Soviet Union's industrial base, it was illegal for foreigners to be owners of Soviet enterprises. The activity of foreign firms was tightly controlled by the central government through the Ministry of Foreign Trade. Foreign companies wishing to buy or sell goods and services to the USSR were required to deal with Soviet foreign trade organizations that served as centralized intermediaries between foreign firms and Soviet enterprises. Because of this centralized control of foreign trade as well as the antagonistic atmosphere created by the Cold War, the number

Reproduced with permission of copyright owner. Further reproduction prohibited.

of American companies involved in business with the USSR was modest. Two notable exceptions were Occidental Chemical and Petroleum, and Pepsi-Cola. Occidental's president, Armand Hammer, had been doing business with the USSR since the 1920s, and Pepsi opened its first plant there in 1974 after signing a trade agreement in 1972.

Economic Reforms under Perestroika—By the mid-1980s, it became evident that the system of centralized planning in the USSR was fraught with problems on a scale too massive and far-reaching to be ignored. Seven basic domestic, social, and economic problems of that period were: a precipitous decline in the rate of economic growth; the low level of effectiveness of the Soviet economy; the need to bring about technological modernization; problems associated with improving the quality of everyday life; societal problems of illegal activity and immorality exposed by the policy of *glasnost*; rigidities of centralized price determination; and centralized allocation of resources and products.[1] The open acknowledgement of these problems was the impetus for the formulation of the economic reforms that constituted Mikhail Gorbachev's policy of *perestroika*. The economic reforms envisioned under *perestroika* had as a major objective the establishment of a market-based demand economy to replace the totally centralized command economy. Nevertheless, this change was to occur in concert with a continuing high degree of central planning.[2]

The two pieces of legislation under *perestroika* that opened the doors to foreign investment in the USSR were the Law on the Soviet State Enterprise and the Law on Joint Ventures, both of which were promulgated in 1987. Article XIX of the Law on the Soviet State Enterprise declared that foreign economic activities of Soviet enterprises were to be an important component of overall economic performance, and that all Soviet enterprises had the right to establish joint ventures directly with foreign partners. The Law on Joint Ventures stipulated the conditions for creating and operating foreign joint ventures in the USSR. Joint ventures were envisioned to be activities between partners engaged in manufacturing or services based on shared ownership and joint management of the enterprise, as well as shared risks and division of profits. Joint ventures with foreign partners were accorded a special legal status by the Soviet government and they enjoyed numerous privileges as well as exemptions from a number of laws applying to domestic Soviet enterprises.

As a result of the unprecedented opportunity to invest in the USSR, the late 1980s witnessed a flurry of activity as Soviets and foreigners eagerly signed joint ventures and other forms of business agreements. Beginning with only 23 foreign joint ventures in 1987, by January 1990 American firms alone had signed more than 140 joint venture agreements in the USSR, the third largest number after Germany and Finland.[3] The four major areas of focus were: the production and sale of computer hardware,

Reproduced with permission of copyright owner. Further reproduction prohibited.

software, and processing services; medical and health care products and services; tourism and entertainment; and light manufacturing. Many of the joint ventures, undertaken on a limited scale, were initiated by small American companies. One-quarter of the ventures had a capitalization of less than 100,000 rubles.[4] In addition, a number of America's largest corporations also signed joint ventures with Soviet partners.[5] Nine of these large firms assessed that, although there was a significant amount of political risk in doing business in the USSR in 1989, the risk could be kept at a manageable level by putting little capital at risk, or by ensuring payment for their goods and services in hard currency.[6] In mid-1992, more than 5,000 enterprises from 65 countries had been registered in the former USSR.[7]

Unfortunately, the benefits expected by the partners were not as great as had been hoped, and the difficulties were far greater than anticipated. In fact, it was estimated that in 1990 fewer than 40 of the 140 registered U.S.-Soviet joint ventures were in active operation.[8] By this time, the problems encountered by American firms operating joint ventures in the USSR were accumulating at a rapid pace.[9] The major problems stemmed from the Soviet infrastructure, particularly the lack of adequate living arrangements for foreigners, office space, and communication networks within the USSR. Other problems originated from shortages of materials, as well as bureaucracy.[10] The difficulties in operating joint ventures had become so great that some analysts were advising firms to evaluate other mechanisms for doing business in the USSR such as licensing and compensation trade.[11] The advantages of these alternative entry strategies were less investment, and repayment in kind as opposed to unconvertible rubles. The scarcity of hard currency remained a problem for all types of endeavors. Even licensing, however, was not without risk. The lack of a comprehensive legal structure for business provided limited protection for foreign licensors. The same risk applied to virtually any type of investment.

The Business Climate in the Former USSR in 1992—The dissolution of the USSR at the end of 1991 brought great uncertainty and instability to all aspects of the economies of the newly independent republics. Uncertainty toward foreign investment was reflected in the attitudes of 1,803 Russians in a survey conducted in the winter of 1992.[12] Those who believed that foreign investment was good for the economy were predominantly under 25 years of age (73%), had a college education (67%), and lived in large cities (70%). While 52 percent believed that foreign investment helped create jobs, 40 percent feared that their government was passing control of the economy to foreign interests. There also were marked differences in the sectors of the economy in which foreign investment was welcomed: 44 percent were against foreign investment in the oil and gas industry, whereas 80 percent favored foreign investment in the consumer sector including pharmaceuticals, clothing, hotels, and restaurants. In addition to their

Reproduced with permission of copyright owner. Further reproduction prohibited.

ambivalence toward foreign investment in general, many Russians took a dim view of some of their fellow citizens who they believed were exploiting foreign business relationships for personal gains such as trips abroad and foreign currency.[13]

In view of these vastly changed circumstances in the environment for foreign investment and other types of business opportunities in the former USSR, we sought to determine the reactions and plans of American businesses with experience or demonstrated interest in the republics. Their assessment may serve as a barometer for other firms with such interests.

The Survey

The purpose of this survey was to determine the current status of American business activity in the former USSR, as well as the prospects for such activity in the future. We designed a questionnaire on this subject and distributed it to eighty businesses located primarily in New England which had been active in the republics of the former Soviet Union, or which had expressed interest in doing business there. New England is a region of the United States where businesses have been particularly active in developing opportunities in the former USSR.[14] Forty-two responses were received, including sixteen telephone interviews in which respondents elaborated upon their answers to the survey. This procedure allowed us to obtain more details to interpret the results and formulate implications. A number of questions lent themselves to multiple responses, which should be noted in interpreting the results.

The 42 firms in the survey ranged from small businesses to Fortune 100 corporations. In order to examine trends on entering and exiting the business arena in the former USSR, we asked respondents to indicate whether they were currently doing business there. Seven firms, in fact, were not, but were contemplating entering the market. The 35 firms that had already established themselves in the former USSR entered the market in the following ways: with a Soviet partner (16), directly (10), through a broker or facilitator (8), or with a Western joint venture partner (4) (multiple responses given). The annual sales volume generated in the former USSR by the firms doing business there was as follows: over $10 million (3 firms), $5 million to $10 million (2), $1 million to $5 million (12) and less than $1 million (18). The length of the companies' business involvement in the former USSR ranged as follows: less than one year (4), 1 to 3 years (13), 3 to 5 years (5), over five years (13).

General Assessment of Business in the Former USSR

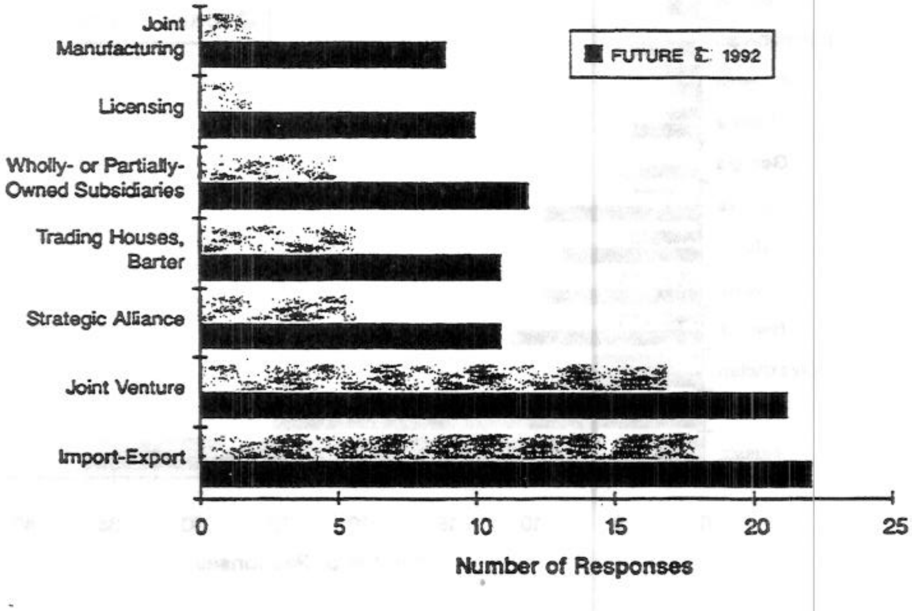
Despite the inherent riskiness of Western business operations in the former Soviet Union, only four respondents evaluated the success of their firms'

Reproduced with permission of copyright owner. Further reproduction prohibited.

endeavors as "poor." The majority (20) rated their firms' success as either "good" or "very good," while eight rated their success as only "fair," and three reported excellent results. These overall evaluations illustrate that the business environment in the former Soviet Union has been relatively favorable for many companies. This is not to say that these results have been achieved without obstacles and risk. This picture does, however, provide support for positive thinking about future operations in these emerging economies.

We asked respondents to describe several aspects of their company's activities in the former USSR at the time of the survey and to indicate how these activities would change in the future. Table 1 shows the types of business relationships the firms had initially engaged in: import-export agreements (18), joint ventures (17), strategic alliances (6), trading houses and bartering (6), wholly or partially owned subsidiaries (5), licensing (2), and joint manufacturing (2). Firms planned to expand their business relationships and recorded increases in all the categories: import-export (22), joint ventures (21), wholly or partially owned subsidiaries (12), strategic alliances (11), trading houses and bartering (11), licensing (10), and joint manufacturing (9).

Table 1. Methods of Conducting Business in the Former USSR

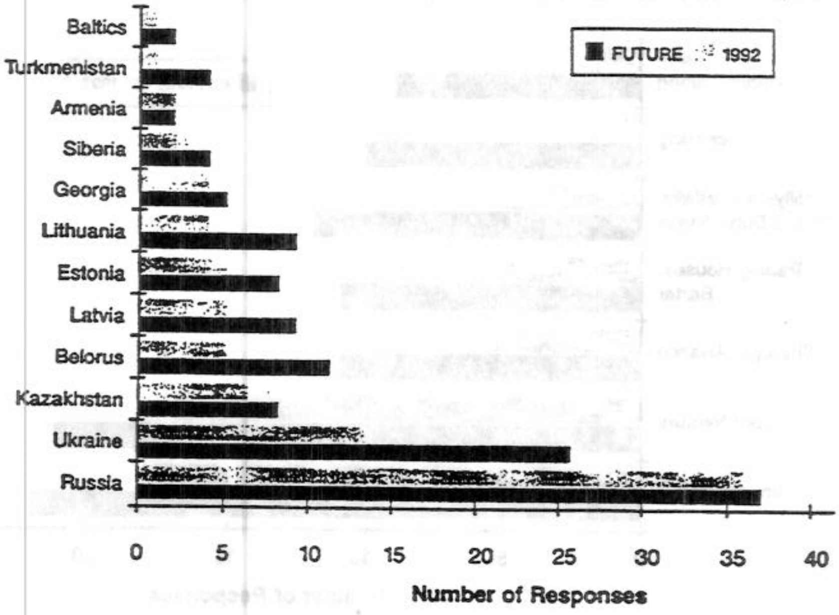


Reproduced with permission of copyright owner. Further reproduction prohibited.

In Table 2 respondents identified the republics in which they were doing business: Russia (35), Ukraine (13), Kazakhstan (6), Belarus (5), and Georgia (4). There was a smaller amount of activity in Siberia (part of Russia), Armenia, and Turkmenistan. In addition, Latvia and Estonia each had five companies operating within their borders, Lithuania had four, while one company simply identified the Baltics as its area of activity. Future plans of these companies showed their intention to maintain their presence in the republics where they had initiated their activities, and to develop a presence in additional republics as well. Areas of future activity were: Russia (36), Ukraine (25), Belarus (11), Latvia and Lithuania (9), Estonia (8), Kazakhstan (8), Georgia (5), Siberia and Turkmenistan (4 each), Armenia (2), and the Baltics generally (2).

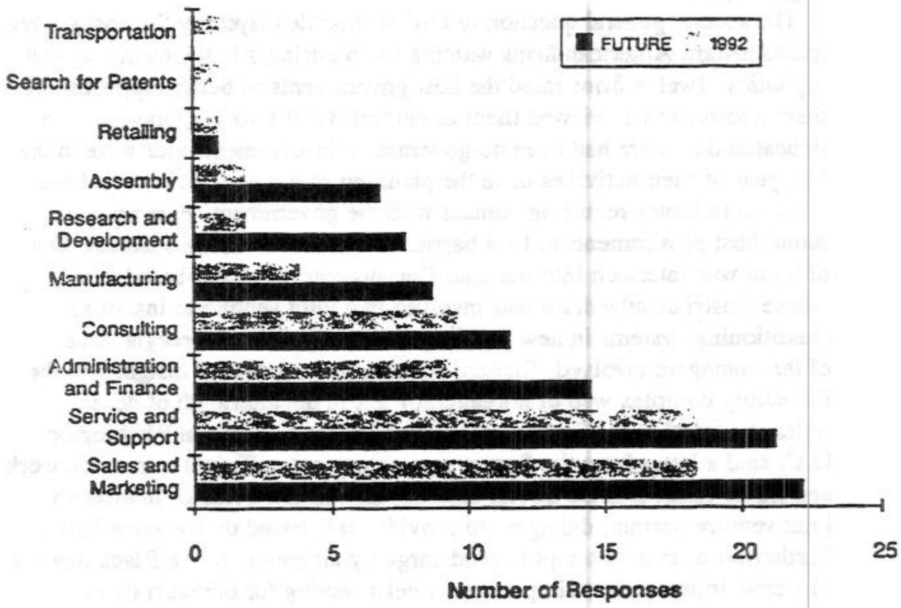
The business functions that the firms performed in the former USSR in 1992 are reported in Table 3 and include: sales and marketing (19), service and support (19), administration and finance (10), consulting (10), manufacturing (4), assembly (2), research and development (2), and retailing, transportation, and search for patents, and an unspecified other business function (1 each). All the functional areas, except retailing, search for patents, and transportation, were targeted for increased activity. Functions with higher levels of activity planned for the future included sales and marketing (23), service and support (22), administration and finance (15),

Table 2. Countries of Business Activities



Reproduced with permission of copyright owner. Further reproduction prohibited.

Table 3. Business Functions Performed in the Former USSR



consulting (12), manufacturing (9), research and development (8), and assembly (7).

Riskiness of the Business Environment in 1992

To get an overall assessment of risk, we asked two general questions. First, respondents rated the uncertainty of the business environment in the former USSR on a 4-point scale. They were nearly unanimous in their agreement that risk was present: Nineteen evaluated conditions as being highly uncertain and highly risky, while 21 judged conditions to be somewhat uncertain and moderately risky. Only two described the environment as having either low or little risk. "Only put in what you can afford to lose. Decide up front what you want—profits, market share, and the like," advised Erich Zarnfaller, Senior International Treasury Analyst at EG&G, Inc., the Wellesley, Massachusetts provider of environmental management services and manufacturer of radiation and security devices. Added Kathy Creculius, a Vice-President at BayBank in Boston: "The former USSR presents . . . great risk. However, for those with ethnic and linguistic ties to the C.I.S. countries, with solid business experience and capital to invest—and risk—there are now opportunities." A recent survey of chief financial

Reproduced with permission of copyright owner. Further reproduction prohibited.

officers of leading companies echoed the sentiments of our respondents regarding risk.[15]

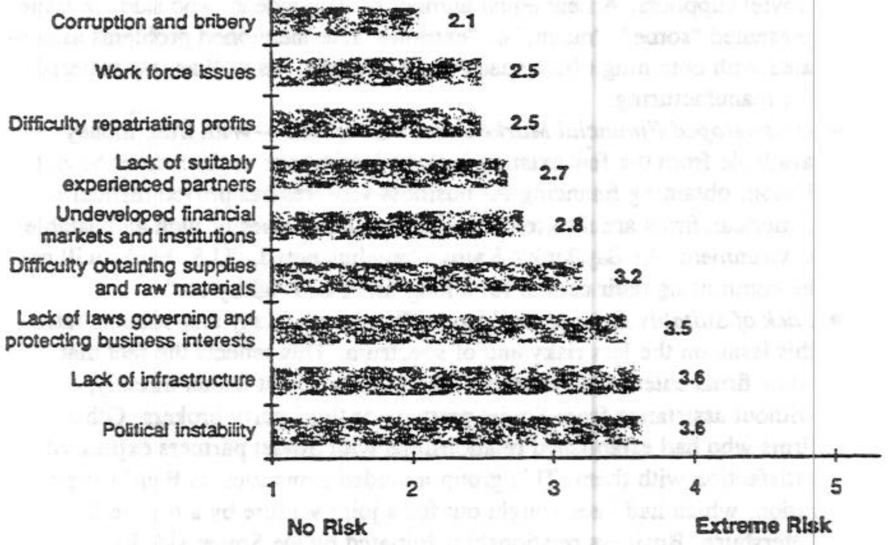
The second general question related to the role played by the host governments toward American firms wanting to do business in the former Soviet republics. Twelve firms rated the host governments as being supportive of their efforts, and 17 viewed them as neutral. Of the six respondents who indicated that there had been no government involvement, four were in the first year of their activities or in the planning stages and had not yet been involved in issues requiring contact with the government. Five firms had found host governments to be a barrier or obstacle to doing business. One of them was Intertech International Corporation, a Boston-based firm whose Austrian subsidiary was involved in a joint venture to install air conditioning systems in new buildings in the republic of Georgia. One of the managers involved, Richard Saint-Amant, provided insight into the incredibly complex web of bureaucracy left in the aftermath of decentralized control in the Soviet Union. The Soviet embassy in Washington, D.C. said a letter from the Soviet partner was not sufficient to provide work and travel visas. But the Soviet embassy in Vienna, home of Intertech's joint venture partner, did agree to provide visas based on the same letter. Furthermore, in an attempt to send cargo by air freight to the Black Sea via Moscow, Intertech lost six precious weeks waiting for bureaucrats in Moscow to "find" the shipment. In that time, the goods could have been delivered by ocean freight to Odessa or air freight to Vienna and trucked to the site.

To obtain a more detailed picture of the risks involved in doing business in the former USSR we had respondents rate nine specific risks on a 5-point scale from "no risk" to "extreme risk." Table 4 displays the means for each category of risk in descending order. The areas of greatest risk were political instability (3.6), lack of infrastructure (3.6), and the lack of laws governing and protecting business interests (3.5). Difficulty in obtaining supplies and raw materials was also rated as a serious risk or obstacle (3.2). Somewhat less serious, but still considered of some risk, were the undeveloped financial markets and institutions (2.8), lack of suitably experienced partners (2.7), difficulty repatriating profits (2.5), and work force issues (2.5). Perhaps surprisingly, corruption and bribery were seen as posing little risk (2.1).

- *Political Instability*—This factor was noted by many respondents as an important source of risk. Nevertheless, Peter Hemingway, Program Director of Russia Operations at Polaroid Corporation in Cambridge, Massachusetts believed that it would pay to be patient: "We still have hopes Russia will emerge as a major trading partner in the international environment. The current problem is lack of focus on where they are going."

Reproduced with permission of copyright owner. Further reproduction prohibited.

Table 4. Risks Faced by Western Businesses



- *Lack of Infrastructure*—Virtually all aspects of the physical infrastructure in the former Soviet Union, from communications to transportation, are inadequate for conducting business efficiently. Henry Quinlan, President of BosMosco, a Boston-based firm marketing food, sporting goods, and medical supplies in Russia, insisted that "lack of distribution presented a tremendous opportunity to create our own distribution system. The entire retail and wholesale network is rising from nothing."
- *Lack of Laws Governing and Protecting Business Interests*—The respondents felt that the lack of laws governing and protecting business interests was a serious risk facing Western firms attempting to do business in the republics of the former Soviet Union. Many laws are in early stages of formulation and the legal framework for the emerging market-based economy is largely uncharted. Western firms have difficulty learning about the laws and preparing action plans. The foundation of laws simply does not exist, and U.S. laws are ignored in that environment. For instance, Lynn Patkin, Vice-President of GreenTree Associates, a computer industry consulting firm in Boston, indicated that some of their clients are encountering difficulty in ensuring compliance with U.S. software copyright laws.
- *Difficulty Obtaining Supplies and Raw Materials*—This issue generated the most disagreement among the people polled in our survey. Many respondents claimed that obtaining supplies and raw materials was not a

Reproduced with permission of copyright owner. Further reproduction prohibited.

problem. These individuals represented service firms involved in banking, consulting, and sales and distribution activities that did not rely on Soviet suppliers. A near equal number of respondents who said the issue presented "some," "much," or "extreme" risk mentioned problems associated with obtaining office space and equipment, as well as raw materials for manufacturing.

- *Undeveloped Financial Markets and Institutions*—With little money available from the few existing lending institutions in the former Soviet Union, obtaining financing for business ventures has proved difficult. American firms are also reluctant to back businesses in such an unstable environment. As BayBank's Kathy Crecelius noted, "U.S. banks will not be committing their capital for a long time, and rightly so."
- *Lack of Suitably Experienced Soviet Partners*—Many respondents rated this issue on the less risky end of spectrum. This reflects the fact that some firms entered the market in the former Soviet Union directly, without assistance from Soviet partners or third-party brokers. Other firms who had established relationships with Soviet partners expressed satisfaction with them. This group included firms such as Baird Corporation, which had been sought out for a joint venture by a firm in St. Petersburg. Business relationships initiated by the Soviet side have become fairly common, as have collaborations involving multiple partners.[16]

  Some of the problems associated with finding appropriate Soviet partners included filtering the opportunities presented by potential partners, and ensuring that individuals are able to back their claims and keep their promises. Ralph Sherman, Corporate Counsel at Innovatech International, Inc., a stainless steel manufacturer in Lexington, Massachusetts, added "filtering opportunities" to the list and rated it as an extreme risk. Ernst & Young, a well-known international consulting and accounting firm, also encountered problems in screening potential partners and opportunities on behalf of several clients. In response to these issues, the Big 6 firm created "Evaluating Soviet Partners," a program designed to identify the advantages and risks of potential projects and partners. Originally intended for the company's internal use, the program was subsequently added to the list of client services due to popular demand, according to John Mahoney at Ernst & Young in Boston.
- *Difficulty Repatriating Profits*—Repatriation of profits did not rank among the highest risks. Both NYPRO Corporation and Intertech stated that the issue was not a problem for them. NYPRO, a Clinton, Massachusetts firm that manufactured moulded plastic products with joint venture partners in Moscow, earned profits in rubles. Conversely, Intertech's Richard Saint-Amant explained: "If somebody really wants to buy, they'll find a way to pay in hard currency." Nevertheless, some respondents

Reproduced with permission of copyright owner. Further reproduction prohibited.

did indicate that repatriation of profits was a major concern for their firms.

- *Work Force Issues*—Work force issues presented little risk according to the respondents. A few, however, ranked the topic as having "much" or "extreme" risk. Some respondents expressed concern about personnel in the host country lacking sufficient understanding of the business principles and practices needed to function effectively in a market economy. Intertech, for example, encountered difficulties trying to employ local Georgian workers. Although Intertech wanted to hire locals and had no problem finding qualified workers, the main contractor they were working for feared problems of poor quality and pilfered supplies, and refused to hire local workers.
- *Corruption and Bribery*—Most firms in the survey had not been exposed to corruption and bribery, a surprising result given media attention to the subject.[17] This is not to say, however, that all firms were free of the problem. One respondent recounted how representatives of his firm encountered police officers in Moscow who stopped them and demanded bribes. They paid the bribe.

Assessment of the Future Business Environment

Investment Plans—Having provided their evaluation of the risks of doing business in the former USSR, the respondents next indicated their firms' investment plans in the region. The majority (69% of firms) planned to either maintain (13 firms) or expand (10 firms) current levels of investment in their existing activities, and 6 additional firms intended to seek out new investment opportunities. The firms who responded with these optimistic intentions included virtually all those with manufacturing operations, around two-thirds of the service businesses, as well as the majority of the sales and distribution companies. Of the remaining 13 respondents, 11 planned to curtail investments, and two, including a small legal firm, had halted business operations completely in the former Soviet Union. Included in the group that planned to reduce their involvement were all respondents from the computer and office products industry. Said one, "With no market determinable, no distribution, and only rubles, how many shareholders will say, 'Hey, that's a great opportunity. Let's do it.'"

Strategic Decisions—We then had respondents forecast how the future business environment would affect their firm's strategic decisions in the former USSR. Respondents checked one of four options. Twenty-six firms were making plans based on the former Soviet Union's development as a strategic future opportunity and expected their presence to increase significantly in size and volume. This group included the majority of firms engaged in all types of activities. Eight firms believed that the former

Reproduced with permission of copyright owner. Further reproduction prohibited.

USSR would be a viable secondary opportunity for them, being minor in size and volume. Three firms, all in sales or services, believed that the region would not be a viable opportunity in the foreseeable future. The remaining five firms, also in sales or services, decided to wait before making any further commitments or decisions.

Respondents also noted the time frame they used to evaluate the business environment. The majority (25 firms) used a fairly short-term window of one to three years. Eleven firms used a slightly longer-term three- to five-year horizon. At the extremes, one firm reported planning over a period of less than one year, and 5 planned more than five years ahead. The executives using the longest time horizon believed that the former Soviet Union would continue to move toward a market economy and planned on its being either a major strategic market or a viable secondary market.

Future Economic Scenarios—To gain an understanding of the basis of the respondents' investment plans, we asked them to indicate which of three scenarios they envisioned as the most likely future direction for the republics of the former USSR. Two-thirds (69%) believed there would be continued movement toward a market economy. Frank Cutitta, President of International Marketing at International Data Group in Framingham, Massachusetts, described the prospects for his firm which publishes the Russian version of *PC World*: "There is a close association between the move to a market economy and the necessity to have the technology to bring manufacturing and production to world levels. We feel we are positioning ourselves to take advantage of this."

With regard to two other possible scenarios for the economy in the former Soviet Union, the remaining respondents (31%) believed that the situation would remain uncertain for the foreseeable future. This group was heavily represented by respondents from the service sector. No respondents in the survey anticipated an immediate movement back to a centrally planned, command economy.

Conclusions and Implications

Despite the obvious risk and uncertainty of conducting business in the countries of the former USSR, our respondents presented a rather optimistic picture of their current operations and their outlook for the future. Two-thirds of firms with operating experience rated their firms' success as either "good," or "very good," or "excellent," and only four described their results as "unsuccessful."

Assessing the Risks—Notwithstanding this positive evaluation of results, respondents recognized the continuing overall level of risk which virtually all described as either "high" or "moderate." The major areas of risk

Reproduced with permission of copyright owner. Further reproduction prohibited.

included the inherent political instability, a grossly undeveloped economic infrastructure, the lack of a suitable legal framework to protect business interests, and the difficulty of obtaining supplies and raw materials. These risk factors were similar to those identified by companies with business experience in the former USSR who responded to a survey conducted by the U.S. Department of Commerce.[18]

Our respondents noted that, within the fragile legal environment, there was much uncertainty about transfer of ownership to private investors, contradictions in the tax code, and undetermined liability protection. Because of the lack of financial markets and institutions, the ability to raise capital was severely curtailed. The unstable ruble was another source of financial risk and uncertainty, and until full convertibility, will likely remain a serious obstacle for Western businesses.

The ability to conduct business in a straightforward manner was seen as being virtually impossible given the serious weaknesses of the economic and physical infrastructure. The lack of reliable suppliers of material and equipment was rated as a severe impediment particularly for manufacturing and assembly businesses, as well as for all businesses needing office equipment and supplies. The communication network was noted as being archaic, overloaded, and unreliable. A related problem was transportation, with roads and railroads being inadequate and in disrepair, and airline travel and transport being extremely undependable.

The overarching risk for Western businesses is clearly the political situation. The unstable and uncertain position of governments at all levels in all countries of the former USSR is abundantly clear to all. The severity of the problem varies among the countries. From civil war in Tadjikistan, Georgia, Ajerbaijian, and Armenia to reactionary forces demanding a return to communist rule in Russia,[19] it is clear that no country is insulated from political turmoil. Economic losses due to repatriation of property by governments, civil disorder, and inability to function in a near-normal business fashion are all realistic scenarios that Western businesses could encounter.

Several issues did not pose major risks for the respondents. The host governments were seen as playing a relatively benign, or even supportive, role. They were seen by relatively few respondents as being an obstacle to their business activities. Work force issues, too, were not seen as an area of notable risk. Neither did corruption nor bribery pose serious problems. A possible explanation for these relatively sanguine views is that the partners in the host countries shielded their Western counterparts from the well-documented complexities and risk in these relationships. The American respondents, after all, had reported little difficulty in locating suitably experienced business partners in the former USSR. Such partners were likely to have cultivated valuable relationships and networks with government officials, suppliers, and competent personnel.

The generally optimistic report from the survey respondents is in stark

Reproduced with permission of copyright owner. Further reproduction prohibited.

contrast to the experience of fifty Russian directors of foreign joint ventures.[20] They reported a deterioration in the situation for joint ventures between mid-1991 and mid-1992, and about half predicted further deterioration over the next six months. They cited the increase of "typically-Soviet" characteristics that interfered with business activity, such as the unclear and often contradictory legislation, the complicated tax structure, incompetent officials, and increasing demands by officials for hard-currency bribes. This is consistent with the economic, legal, and other problems associated with the collapse of the Soviet ministries.[21] The recognition of these mounting problems by senior Russian joint venture directors reinforces the need for Western businesses to develop relationships with experienced and reliable partners.

Reason for Optimism—Although our respondents, as noted earlier, realized that this transitional environment was inherently risky, they remained optimistic about the future of the former USSR. Sixty-nine percent intended to at least maintain, or even expand, their activities there, and a large majority believed that the countries would continue to move toward a market economy. Virtually all saw these countries as continuing to be a significant strategic opportunity for their businesses or, at least, a viable secondary market. Our respondents did not see repatriating profits to be a major risk. The reason for this might be the lack of hard-currency profits being realized at the time, or planned in the foreseeable future. This overall optimistic viewpoint underscores the attitude toward risk that Western participants in the former USSR must maintain.

Despite their willingness to accept uncertainty and risk, the majority utilized a one- to three-year time frame for their decision making. Although this may seem a relatively short time period, the uncertainty itself would likely prevent most executives from looking further into the future. Clearly, an attitude of fundamental entrepreneurialism and risk taking is the basic requirement for doing business in this arena.

Consistent with their optimistic views about the future, most respondents intended to expand their activities both geographically and functionally within the countries of the former USSR. They planned to expand primarily in the European countries of Russia, Ukraine, Belarus, and the Baltics. These countries have business conditions most similar to those in the West, and constitute potentially major markets for goods and services. The remaining countries are less prepared to accommodate Western-type business activities, and present greater risk, although Kazakhstan was seen by a substantial number of firms as an area of significant opportunity. Additionally, a number of respondents planned to increase their activities in sales and marketing, service and support, manufacturing and assembly, R&D, and consulting. As they expand, they would be well advised to work closely with their partners both to benefit from local expertise and relationships,

Reproduced with permission of copyright owner. Further reproduction prohibited.

as well as to help in developing their partners' abilities in these new functions.

Rethinking the Form of Business Activities—Despite the prevailing optimism and the reported success of some joint ventures such as Dialogue,[22] it is clear that some companies are rethinking joint ventures as the predominant form of entering the former USSR. The respondents to our survey indicated slightly less of an increase in this form for the future, compared to others such as licensing, joint manufacturing, wholly- or partially-owned subsidiaries, strategic alliances, and trading houses or bartering. There are a number of possible explanations for exploring alternatives to joint ventures. First, they are the most complex form of cooperation since they require difficult negotiations in the formulation stage, as well as joint decision making, power sharing, and profit sharing during the operational stage. Second, the benefits initially granted by the Soviet government to joint ventures have been greatly reduced since the dissolution of the Soviet Union. Joint ventures no longer enjoy tax breaks and other benefits and exemptions from Soviet laws that once made them attractive. Since new legislation has yet to be clarified in the independent republics, foreign companies wanting to establish joint ventures need to protect themselves by getting approval from as many authorities at as many levels as possible. Finally, risk often accompanies joint venture activities in the form of higher levels of investment than is required with some other forms of business activity.

Experience with joint ventures has, in fact, been mixed. NYPRO president, Gordon Lankton, was bullish on his joint venture: "Opportunities are exceptional at present. [The former Soviet Union] is a great place for joint ventures." His plastic injection moulding plant, Rotel-Mikma, generated ruble profits in 1992 and the firm had plans to increase investment and install state-of-the-art injection molding equipment. NYPRO had the good fortune of joining an already established joint venture with reliable Russian and Swiss partners. The experience of Data General Corporation was not so positive. The Westborough, Massachusetts computer manufacturer had entered the Soviet market in a joint venture with the Soviet Ministry of Electronic Automation. According to Bill Burek, Assistant General Counsel for International Operations, the firm considered the former Soviet Union to be a significant future market, but DG's venture had not achieved the success that had been anticipated. When the communist regime collapsed at the end of 1991, the ministry was dissolved, leaving no Soviet participants in the joint venture with the authority to commit resources for compensation on transactions. In short, the window of opportunity closed. For the short term, Data General planned to curtail its investment in the joint venture, but planned to strengthen its commitment in the longer term once the political and economic situation showed signs of stabilizing.

Reproduced with permission of copyright owner. Further reproduction prohibited.

In recognition of the difficulties and constraints associated with joint ventures, more of our respondents planned to increase other types of business activity. Other methods of dealing with such difficulties were suggested by American executives in a United States Department of Commerce report.[23] These executives, all with experience in the former USSR, recommended actions that the governments of the former Soviet republics, as well as the American government, could take to attract investors and reduce the risks of doing business in the former USSR.

Surviving the Roller Coaster Ride

The generally promising, but still complex, situation is illustrated well by the observation of one high-tech executive in our survey. "The Russians are confused by what's going on. They are cut off from supplies and communications. You can't really call them businessmen—they have no experience. But I was extremely impressed with the idealism, intelligence, and excitement of the new young workers."

It is clear, then, that Western firms must continue to carefully evaluate the risks and rewards in the countries of the former Soviet Union. "We are looking at the market from a long-term viewpoint. We feel it's important to ride the roller coaster in order to be there," commented Frank Cutitta of International Data Group. Those who would operate in this environment must be prepared for the bumpy ride and constant twists and turns. The trip will require perseverance and a long-term outlook, but some observers suggested the possibility of a smoother ride after Yeltsin's decisive action against the reactionary Russian parliament in October 1993.[24]

References

1. I. Faminsky and A. Naumov, "Historical Review of U.S. and Soviet Economic Institutions," in P.R. Lawrence et al., *Behind the Factory Walls: Decision Making in Soviet and U.S. Enterprises* (Boston, MA: Harvard Business School Press, 1990), pp. 37-55.
2. J.S. Berliner, "Soviet Management from Stalin to Gorbachev: A Comparison of the Harvard Project and SIP Interviews," in J.S. Berliner, ed., *Soviet Industry from Stalin to Gorbachev: Essays on Management and Innovation* (Ithaca, NY: Cornell University Press, 1989).
3. "Strangled by Red Tape," *The Economist*, February 24, 1990, p. 72.
4. D. Nigh, P. Walters, and J.A. Kuhlman, "US-USSR Joint Ventures: An Examination of the Early Entrants," *Columbia Journal of World Business* (Winter 1990), pp. 20-27.
5. D. Nigh and K.D. Smith, "The New U.S. Joint Ventures in the USSR: Assessment and Management of Political Risk," *Columbia Journal of World Business* (Summer 1989), pp. 39-44; N. Giges, "Colgate, P&G Pack for Road to Russia," *Advertising Age*, March 12, 1990, p. 56; G. Rifkin, "I.B.M. and American Air in Aeroflot Deal," *New York Times*, April 23, 1992, pp. D1, D6.
6. Nigh and Smith, op. cit.
7. P. Razin, "Hard Times for Joint Venture Directors," *Delovie Lyudi* (July/August 1992), pp. 38-39.

Reproduced with permission of copyright owner. Further reproduction prohibited.

8. K.A. Rosten, "Soviet-U.S. Joint Ventures: Pioneers on a New Frontier," *California Management Review*, 33/2 (Winter 1990): 88-108.
9. J. Kaufman, "Reality Hits U.S. Firms in Soviet Union," *Boston Globe*, June 17, 1990, pp. A1, A5; R. Neubecker, "Big Deals Run into Big Trouble in the Soviet Union,: *Business Week*, March 19, 1990, pp. 58-60.
10. Rosten, op. cit.
11. Ibid.
12. B. Weber, "Many in the Former Soviet Lands Say They Feel Even More Insecure Now," *The New York Times*, April 23, 1992, p. A3.
13. D. Kramer, *The Business Environment Since the Fall of the USSR*, presentation at the England Slavic Association meetings, Russian Research Center, Harvard University, April 10, 1992.
14. K. Blanton, "Mass Firms Look to the East," *The Boston Sunday Globe*, September 15, 1991, pp. A37, A39.
15. J.B. Welch, "Investing in Eastern Europe: A Survey of Chief Financial Officers," *The International Executive*, 35/1 (1993): 45-72.
16. A.B. Sherr and M. Hayes, "Joint Venture Partners Speak Out: Interviews and Surveys of Participants," *East European/Soviet Union Executive Guide*, June 3, 1991, pp. 18-21.
17. F. Kaplan, "Corruption Overwhelms Russia's Top Lawman," *The Boston Globe*, September 27, 1992, pp. 1, 19.
18. U.S. Department of Commerce, *Obstacles to Trade and Investment in the New Republics of the Former Soviet Union* (Springfield, VA: International Trade Adminstration, U.S. Department of Commerce, National Technical Information Service, March 1992).
19. F. Kaplan, "With Yeltsin's Foes Disbanded, Difficult Path to Reform is Reopened," *The Boston Sunday Globe*, October 10, 1993, p. 2.
20. Razin, op. cit.
21. A. Shama and S. Sementsov, "The Collapse of the Soviet Ministries: Economic and Legal Transformation," *The International Executive*, 34/2 (1992): 131-150.
22. R. Brady and D. Greising, "They Know How to Capitalize on Chaos," *Business Week*, April 27, 1992, pp. 106, 107.
23. U.S. Department of Commerce, op. cit.
24. M. Vaillancourt, "Moscow: Business as Usual; Stronger Yeltsin Regime Seen as a Stabilizing Factor," *The Boston Globe*, October 6, 1993.

Reproduced with permission of copyright owner. Further reproduction prohibited.

BEST ORIGINAL AVAILABLE

5A

Daniel J. McCarthy and
Sheila M. Puffer

Columbia Journal
of World Business
v30n3 Fall'95

Diamonds and Rust on Russia's Road to Privatization: The Profits and Pitfalls for Western Managers

LIPPINCOTT LIBRARY
DOCUMENT DELIVERY

McCarthy and Puffer focus on large, state-owned Russian enterprises, and explore what changes have occurred in them during the recent privatization process. They compare the conditions facing Russian managers in these firms at the beginning of 1995 with the situations reported by managers of state–owned enterprises during the early stages of privatization in 1992. They make several recommendations to Western business managers to assist them in uncovering the potential "diamonds" among Russian enterprises, rather than wasting time fruitlessly scraping at the rust of others.

The second phase of privatization in Russia officially began on July 1, 1994, with the major objective of attracting new capital to enterprises. This phase followed the voucher privatization phase which aimed at transferring ownership of enterprises into private hands without an infusion of additional capital. By many accounts, "Phase One" succeeded in accomplishing this limited objective, while leaving the condition of many enterprises unchanged, yet providing a potential platform for real change in others.

The current status of privatization in Russia is important to Western business managers who currently are, or plan to be, doing business in Russia. In joining the worldwide trend toward privatization,[1] many Russian enterprises have been moving toward independence from central ministries and planning agencies since the disintegration of the Soviet Union.

The analogy of "diamonds and rust" refers to the promising and stagnant enterprises in Russia's privatizing economy. Many new private start–ups, as well as other small and medium–sized organizations that were privatized by 1994, could be described as diamonds in the rough. Each shows the promise to succeed in a market economy. Other firms, either privatized or not, still operate essentially as state–owned enterprises, having undergone little or no real change. The managers of the latter organizations have operated essentially as they did in the pre–perestroika command economy, and could be called the rust of the system.

The Diamonds

Some larger partially–privatized or still state–owned enterprises such as Gazprom, LukOil, and aerospace firms such as Mikoyan and Sukhoi,[2] possess resources, entrepreneurial leadership, and ready markets for products and services. Other potentially powerful and profitable corporate giants are in the mining, nonferrous metals, electricity, and telecommunications sectors.[3]

Shares of LukOil, the world's largest oil producer, for instance, found favor with Western investors like Credit Suisse First Boston which bought 3% of its shares during voucher privatization, the first phase of the Russian government's privatization program.[4] Gazprom, with revenues of over $11 billion from its export network throughout Eastern and Western Europe, is reputed to be worth between $250 and $740 billion if its reserves were valued like those of Western energy companies.[5] Yet, given its present understated market value in Russia, its capitalization is a mere $250 million. Gazprom is obliged to sell 9% of its shares for cash within the second phase of the privatization program, and the Russian government plans to sell shares it owns in a global offering.[6] Although Gazprom, with its 320,000 employees, is still somewhat of a mystery to many analysts, it has been called by one Western energy executive "the most professional of all the Russian companies with whom he had dealings."[7]

Along with many smaller entrepreneurial companies, enterprises such as Gazprom represent the diamonds which attracted an estimated $3 billion in Western investment during 1994. More impressive is the esti-

Daniel J. McCarthy is the Philip R. McDonald Professor of Business Administration in the College of Business Administration at Northeastern University. He is the author or co-author of ten articles on Russian management as well as the book, Business Policy and Strategy: Concepts and Readings. *Sheila M. Puffer is an Associate Professor in the College of Business Administration at Northeastern University and a Fellow at the Russian Research Center at Harvard University. Recent publications include* The Russian Management Revolution: Preparing Managers for the Market Economy *(1992) and* Management and Business in Russia *(forthcoming).*

mate by a senior manager of a large U.S. investment house who forecasts an infusion of $20–$25 billion within six months, if appropriate legislation were passed.[8]

The Rust

Despite the potential of such viable enterprises, many other organizations, particularly very large partially–privatized or still state–owned enterprises, are encased in 70 years of rust, will take years to polish, and many will eventually end up as scrap. In fact, their situation had worsened by 1995, in step with the perilous decline of the Russian economy. The 1995 state budget, for instance, called for "tight spending limits and cutting credits to failing state enterprises."[9] Prime Minister Chernomyrdin evidently "accepted the International Monetary Fund's view that there is no alternative to tough fiscal and monetary policies."[10] Such actions can only lessen the survival chances of many large state–supported enterprises.

Some enterprises like Zil, a major truck and automobile conglomerate, should have been declared bankrupt; but such action would have put the future of many thousands of employees in jeopardy. The new competitive environment eroded the market for Zil's products, and the situation was aggravated by management's inability to initiate change, as well as the enormous cost structure of a bloated work force, and the failure to attract enough capital for retooling to establish a competitive position. These circumstances converged to undermine a company which had been touted as a model enterprise within the former centrally planned, monopolistic system. A former jewel had thus become corroded in the emerging competitive market system, as its managers were unwilling or unable to effect meaningful change.

Privatization's First Phase

July 1994 marked the end of the first phase of privatization, which involved the distribution of vouchers to Russian citizens, conveying some of the ownership of many state enterprises into private hands, predominantly workers and managers. By many accounts, this phase successfully achieved its objectives. Over 19,000 of 29,000 designated enterprises had been privatized,[11] over 40 million Russian citizens owned stock,[12] and more than 60% of the labor force was employed by the private sector, specifically by enterprises in which the state was not the major shareholder.[13] Even the most hardened skeptics were surprised by the pace and scope of the privatization effort. The objective of placing some ownership of enterprises into private hands had been accomplished, although the state was still a major shareholder of many such partially–privatized companies.

In fact, much of what has been written about entrepreneurship and competition in Russia applies to smaller privatized companies, usually in the retail and service sectors, as well as private start–up organizations. Also included would be some smaller manufacturing companies that were mostly start-ups rather than privatized versions of former state–owned enterprises. Still, of the smaller and medium–sized organizations that were formerly state–owned, virtually all were at least partially in private hands by the end of 1993.[14]

Many of these smaller enterprises adapted somewhat successfully to the emerging market economy for several reasons. They required relatively little capital, they were not burdened with large work forces, were closer to customers than their larger counterparts, and were very flexible in response to changing conditions. Also, because many of these organizations operated primarily on a cash basis, they could avoid paying excessive taxes, as well as escape the problem of nonpayment on their receivables that in mid-1994 totalled $10 billion nationally.[15]

New ownership, however, did not necessarily change the nature of many other organizations, nor prepare them to function in a market economy. For many, the state remained a powerful owner, and for others the vast majority of shareholders were the same managers and workers who had operated the overstaffed and ineffective state enterprises. The objectives of these new owners had little to do with generating profits for reinvestment in their enterprises to help them become competitive entities. Rather, workers were interested in job security, as were the managers whose positions often depended upon satisfying workers' demands. Some analysts have reported that more than 70% of the shares of privatized companies ended up in the hands of such workers and managers.[16]

The Second Phase of Privatization

Phase Two of privatization began in mid–1994 with the objective of infusing new capital into Russian enterprises. The conversion of these organizations to a condition where they would be able to truly operate as private enterprises should require an enormous amount of investment capital which was unavailable during Phase One. Unless the second phase of privatization is successful, it is unrealistic to expect major changes in most large enterprises owned partially or wholly by the state. It is still questionable whether a major infusion of capital will foster real change in those organizations which were unable to adapt to the first phase of privatization. Yet, the success of these organizations could be vital to employment, social stability, and the continued development of a market economy in Russia.

Although Phase Two of privatization began officially on July 1, 1994, few enterprises had actively engaged in the process by the end of that year. The primary reason was a conflict between the objectives of enterprise managers and those of the state, both parties being powerful shareholders. The state, while continuing to grant subsidies, sought financial results, whereas managers sought to remain in power.[17] State enterprise managers have been described as being internally focused, looking to the past, resistant to change, and closeminded and slow to act on new ideas.[18] Thus, in mid–1994, in order to make the remaining state–owned enterprises more attractive to investors, Anatolii Chubais, then the head of the privatization effort, set a limit of 10% share ownership for workers and managers, rather than the 51% allowed under Phase One.[19] Clearly, he was trying to prevent the buildup of even more rust in these enterprises.

It was recognized by many, in fact, that little had changed in most enterprises in an extremely important sector of the Russian economy, namely very large organizations, 70% of which had been declared privatized by mid–1994.[20] Whether privatized or not, most were still operating at the beginning of 1995 essentially the way state–owned enterprises had for decades. For instance, the enormous monopolies of the defense sector still employed countless thousands of workers, had no real demand for their products, produced for inventory to maintain employment, were often not paid even by the government for products they delivered, were unable to invest in new plant and technology, and still relied upon government subsidies to continue operations. In fact, lobbies for the defense, heavy industry, and agricultural sectors obtained sizeable increases in government credits during 1994.[21] One analyst estimated the total value of credits during the summer of 1994 to be greater than 13 trillion rubles,[22] and the otherwise austere 1995 draft budget still called for maintaining defense spending at 5.1% of GDP.[23] In a mid–1994 survey of 158 defense companies, only 65 of which had been privatized, 84% reported receiving preferential government credits.[24]

Even strong and well positioned companies like Gazprom, the natural gas giant, had reportedly changed very little. "Gazprom has an aura of success about it," according to a Moscow–based investment banker. "But if you look close, it's a typical Russian organization. No one on the inside understands what the financial condition of the company is."[25] One financial analyst stated that it could take ten years to audit the enterprise.[26]

Yet, even in the very difficult environment for newly privatized enterprises, some managers have attempted to scrape the rust from their former state–owned enterprises. A case in point is Joseph Bakaleynik, a Harvard MBA who returned to the Vladimir Tractor Factory in May 1994 and was elected general director by the worker–manager stockholders. Inheriting a $3.5 million debt and anticipating few new orders, he immediately trimmed the 16,000–person work force by a third, cut back extensive employee social services, and imposed tighter accounting controls. To gain hard currency, he focused on the export market because 21% of the company's 1993 tractor output had come from sales to the U.S., Canada, and Mexico.[27] With such leadership, this enterprise could well be a diamond in the rough.

Although privatization's first phase thus seemed to have achieved its primary objective of large–scale private ownership of the Russian economy, some would argue that what had occurred was not privatization but "de–statization." One conservative Russian analyst noted that while, as of January 1994, 56% of Russian enterprises had been officially privatized, only 7% of the reported industrial production of those enterprises was in fully privatized firms. Forty–three percent of industrial production, in fact, came from enterprises where the state was still the sole owner, and an additional 39% came from firms where the state continued to own a significant share.[28]

These data support the dominant role of fully state–owned and predominantly state–owned enterprises in the basic Russian economy, and explains in part why so little change occurred in many such organizations during the first four years of Russia's privatization effort. This paralysis resulted, among other things, in a decline in industrial production of over 22% for the first ten months of 1994 compared to the same period for 1993.[29] The task of turning around many large enterprises during Phase Two of privatization is formidable indeed.

Comparison of Enterprises in 1992 and 1995

The responses of 57 managers of state–owned enterprises to a survey conducted by the authors in 1992 can serve as a comparison to the situation faced by enterprise managers as they entered 1995. In the survey, managers expressed their views on the challenges and opportunities they encountered, as well as their personal and business objectives as their enterprises moved toward privatization. In addition, some managers provided elaborations of their responses during interviews. Describing the sources of the problems they faced, the managers included government regulations and the political situation, suppliers and customers, labor, financing, technology, and competition. They also described the importance of factors contributing to the success of their enterprises as indicated by sales growth, profitability, personal fulfillment, and employee development. Presented below is a comparison of the situation they described in 1992 with the situation at the outset of 1995.[30]

In 1992, the environment in which the managers of the state–owned enterprises functioned was characterized by turbulence, uncertainty, and change. As shown in Table 1, the most severe problem was obtaining a reliable source of supply. Government regulations, a highly unstable political situation, difficulty obtaining financing, and a lack of advanced technology were other serious problems. Marketing, competition, and customer demand, as well as availability of qualified labor, were not seen as major obstacles.

Entering 1995, many such managers were operating partially privatized organizations, yet still faced the same problems which were considered most important in 1992, as well as additional ones. As in the past, government regulations were still confusing and unevenly enforced, and the political situation, although improved, was still very unstable. In fact, one observer reported that the fabric of Russia's political stability was unravelling.[31] Nonpayment of debts between companies, and even by the government, had put many companies into virtual bankruptcy. Yet, the government was slow to close them, and doled out inadequate subsidies which were used to retain oversized, but often unpaid work forces, but were not substantial enough to upgrade plant, equipment and technology. Sources of supply were still highly unreliable, exacerbated by the chronic nonpayment of inter-firm debt within the supply chain. Unless Phase Two of privatization could bring in desperately

Table 1

Sources of Problems for State-Owned Enterprises in 1992 and 1995

Note: The 1992 survey reports responses of 57 managers, including 33 directors and deputy directors. Enterprises were reported as state-owned by 70 percent of respondents, 14 percent were leased from state enterprises, and 16 percent were privatized stock companies. Enterprise activities included manufacturing of industrial products (34%) or consumer products (13%), services (19%), R&D (21%), and other (13%).

Problem	*1992 Survey Results Mean & Interpretation (7-point scale)*	*1995 Estimates*
1. Suppliers	5.3, serious problem	still serious
2. Government regulations	5.1, serious problem	still serious
3. Financing	5.1, serious problem	more serious
4. Political situation	4.8, quite serious	more serious
5. Technology	4.6, quite serious	more serious
6. Customers	3.9, somewhat serious	very serious
7. Labor	3.8, somewhat serious	more serious
8. Marketing	3.3, slightly serious	very serious
9. Competition	3.0, slightly serious	very serious

needed capital, the situation would remain extremely difficult.

Other problems noted as less serious in 1992, such as marketing, competition and customer demand, had become much more acute by 1995. The lack of marketing had become a very serious weakness as competition became severe and customer preferences shifted to imported goods. Imported goods constituting 50% of goods sold to consumers at the end of 1994, according to some reports.[32] Among the managers we interviewed in the fall of 1994 who suffered from stiff foreign competition was the general manager of Mikromashina, a formerly successful state–owned producer of good quality electrical consumer products. He had not yet found a way of successfully marketing his goods to Russian consumers who now preferred to buy imported rather than domestic products. Severe cash problems and plummeting demand forced him to drastically reduce the work force and rent some of the facilities to other enterprises.

The objectives reported by the managers surveyed in 1992 represented a relatively balanced set of personal, business, and financial goals including breaking even and growing sales, while striving for personal independence and fulfillment. All of these goals were seen as relatively equal in importance. The managers also noted six factors that were important to business success, again with no appreciable differences among them. These were skills and attitudes of employees, quality and uniqueness of the product or service, profitability of the business, level of technology, and competitive price. The managers seemed to understand that success required diverse enterprise attributes, including good products, employees, and financial situations. Again, marketing received little attention. As for strengths of their enterprises, managers in 1992 reported strong demand, a capable work force, and product or service uniqueness. Closely following were a strong financial position, managers' own abilities, and a positive company image. Other factors, including advanced technology, export capability, and good marketing, were not considered their organizational strengths.

As noted earlier, at the beginning of 1995 privatization was six months into its second phase, yet few former state–owned enterprises had successfully adapted to the new circumstances or attracted new capital. Even though in 1992, managers had established fairly realistic objectives, seeming to understand their enterprise strengths and what it would take to succeed, most would apparently be unable to effect the changes needed to do so. As external problems worsened through the first half of the 1990s, enterprise strengths deteriorated, and weaknesses like those reported in 1992 prevented progress in the marketplace. As a result, managerial objectives focused on survival. Senior managers seemed preoccupied with struggles for ownership and control, concentrated on external issues and activities, and distanced themselves from internal operations. While first–level line managers and workers sought job and wage security, senior managers sought to secure their positions by consolidating ownership, establishing parallel commercial and financial activities, re–integrating their enterprises into state economic and political structures, and re–establishing old relationships in the production sphere. Making matters worse, many of the best qualified employees left for higher paying jobs and better career opportunities in private start–ups, creating a shortage of motivated and qualified workers.[33]

Essentially, by 1995 even managers of organizations that in the early 1990s were well positioned to succeed under privatization had reverted to the former state–owned enterprise way of managing and operating.[34] The divergent objectives of senior managers, and those of line managers and workers, impeded progress in reorganization, rationalization of the work force, and other meaningful methods of cutting costs, thus preventing increased productivity and improvements in product quality and marketability. A lack of marketing skills and intense competition from foreign products also contributed to a marked decrease in demand and, ultimately, worsening financial situations. Such was the situation for many partially–privatized and state–owned enterprises at the outset of 1995, a sharp deterioration from the more promising scenario of 1992.

Nonetheless, by mid-1995, many positive developments had occurred in the political and economic environment. These developments, discussed in the article's conclusion, should provide a much more positive atmosphere and create many opportunities for those enterprises which are able to adapt the strategies and cultures of their organizations to these improving circumstances.

Summary: Diamonds in the Rust?

The first four years of privatization provided an opportunity for many enterprises to shed the shackles of total state ownership and centralized direction. Some made progress because of superior resources, advantageous positions in the economy, or the ability to export for hard currency. Like some private start–ups, these firms adapted well to a market economy, and established competitive positions. Such organizations grasped the opportunity to prosper, and their luster could be attractive to Western investors and partners.

Not all privatized organizations, however, would be able to rid themselves of rust. The Russian managers surveyed in 1992 understood, on the surface at least, the severity of the problems accompanying privatization. It was uncertain then who would own the enterprises, what the roles would be for the former power brokers of the discredited centralized system, what the nature of government regulations would be, and whether the political situation would stabilize. Unfortunately, most managers had little control over these and other critical elements of their environment, which deteriorated as privatization progressed.

Nevertheless, managers of some powerful state–owned and partially–privatized enterprises lobbied hard for their own interests, including a slowing of the privatization process. Their actions, as well as those of the Russian government, effectively sidetracked real change, preventing effective adaptation of many such enterprises to the emerging market economy. In fact, even at the outset of 1995, few enterprises had actively participated in privatization's second phase of attracting new investors due to the delaying tactics employed by government officials and managers who sought to establish their respective positions of power and control. Many enterprises continued to function as monopolies receiving state subsidies, and employed the failed policies which existed before privatization.

This situation occurred in part because of attitudes developed during the shortage economy in which these enterprises had previously operated. The managers in 1992, for instance, had expressed relatively little concern for the importance of marketing as either a source of problems or a company attribute, and experienced ready demand for their products. Also, many seemed constrained by the expectation of keeping their work forces employed regardless of the resulting inefficiencies. Finally, their break–even objective might have prevented them from taking the difficult turnaround actions necessary to generate funds for investment in updated plant, equipment, and R&D.

Thus, if real progress were to be made, as privatization progressed, many things should have changed in Russian enterprises. It would, for instance, have been crucial that managers learn how to link their privatization strategies with their business strategies in the interest of long–term profits.[35] Some responses by managers in 1992 did give indications that positive changes had begun, and in the subsequent two years, some enterprises and managers did adapt to the requirements of the market economy. Still, success in the difficult conditions faced by most Russian managers and enterprises would depend upon successful implementation of privatization's second phase, which in turn could depend upon the involvement of Western managers and investors.

Profits and Pitfalls for Western Managers: Recommendations for Action

In spite of the problems that have persisted under privatization, many opportunities exist for Western firms to discover gems within the rusted economy of Russia, and a sizeable number have already found them.[36] Honeywell, for example, exceeded its 1993 annual sales targets by March, and Hewlett Packard recorded a 400 percent increase in its Russian sales. Motorola, Reebok, McDonald's, Philip Morris, Coca–Cola, Pratt & Whitney, IBM, and ABB Asea Brown Boveri also figure among the successful prospectors.[37] Although many opportunities will be sought with new private enterprises, managers of Western firms will also inevitably work with those in state–owned or partially–privatized enterprises either as customers, suppliers, joint venture partners, or in other strategic alliances.

The following recommendations are aimed at assisting Western managers as they develop mutually beneficial relationships with state–owned or formerly state–owned organizations. Examples are provided of how a number of firms have already overcome obstacles and taken advantage of the strengths that some Russian enterprises have to offer. In spite of the pitfalls which still exist under privatization, there is profit to be realized from doing business in Russia.

1 *Help Russian managers resist the temptation to backslide to old ways.*

Entering 1995, many state–owned and partially–privatized enterprises were in a survival mode, just as in

1992 when respondents stated that breaking even was their primary objective. Managers faced the threat of having their enterprises declared bankrupt and put up for auction if they did not pay their debts within a three–month period. Furthermore, the top management was to be replaced during this process, a compelling reason to meet the break–even goal.[38] In fact, far fewer enterprises were liquidated than expected, thanks to government subsidies, even through mid–1994. This period, though, was to mark the end of government largesse, as described by one analyst who stated, "after the big bailout settles debts, enterprises are supposed to go on a strict bottom–line basis. If they can't make it then, the new bankruptcy law is supposed to go into effect and their assets will be sold off."[39]

On June 15, the government estimated that over 1100 firms would be declared unprofitable and be reorganized or sold. Profits for January 1994, in fact, had dropped to approximately one–third of those a year earlier.[40] Yet, by late 1994, only four state–owned enterprises were involved in bankruptcy proceedings. One of them, Kristall Distillery, the producer of Stolichnaia vodka, was declared insolvent by the Federal Bankruptcy Administration. Yet, the company seemed to be successfully maneuvering to avoid final bankruptcy, which the government also wanted to avoid.[41] Clearly, little sustained pressure has been applied to state–owned enterprises and managers in spite of government policies to the contrary.

It is not surprising, then, that backsliding to old ways became a major strategy for many managers of state–owned and partially–privatized enterprises. Survey responses in 1992 about enterprise strengths suggest why this happened. Viewing a strong demand for their products and services as a major strength of their enterprises could well have led to complacency, and not recognizing competition as a serious problem probably perpetuated monopolistic practices. The managers of a Smolensk bakery and candy manufacturer, for instance, acquired privatized status in early 1993. Although the staff gained ownership of 61 percent of the shares, virtually everything remained as it had been, even the large mural of Lenin behind the director's desk.[42] Similarly, the newly privatized Smolensk Flax Kombinat, probably the largest linen manufacturer in the world, continued to produce and sell the same Stalinesque chintz designs they had made for years.[43] An explanation for such resistance to change may be a reluctance to assume responsibility and take initiative, risk–averse behaviors that had a long history in Russia and were reinforced by the centrally planned communist system. Even when some managers tried to make changes in their companies, entrenched forces and habits prevented them from doing so.[44]

Western managers must recognize these tendencies and provide the controls and measurements to hold Russian managers accountable for their performance. For instance, the Krasnyi Proletarii machine tool plant languished for two years without orders until it radically restructured its operations, cut the labor force by a third, developed a rational business plan, and involved employees at all levels in decision making. They accomplished this turnaround with the assistance of an American business consultant who pressed them to initiate painful but mandatory changes that required them to take responsibility for the company's action and results. The company was later viewed as a model for other state enterprises, partly because it generated $1 million in export sales of its good quality, well-priced products at a Chicago trade fair in 1993.[45]

2 *Recognize and deal with political and legal uncertainties.*

The most serious obstacles to emerging gems in the Russian economy have been capricious and contradictory government regulations and the highly unstable political situation. In 1992 these were the major problems for state–owned enterprise managers, and in 1995 still interfered with managerial decision making. Western managers need reliable and effective business partners to navigate through the confusing legal and political environment.

Uncertainty and ambiguity have made it very difficult for Russian managers to operate in the short term, and nearly impossible to plan for the longer term. In this environment, a 1992 survey participant, the general manager of the Domodedogo Sewing Factory, expressed a highly conservative approach to his business: "First of all, in a world where we all now live, in a very complex world, about two or three years ago we had to set for ourselves the objective to survive. During this transitional period when old relationships have been destroyed and new ones are being established with great difficulty, our goal is to preserve the organization and production which we have today."

Western managers, too, have experienced the difficulty of operating in this unstable environment. The oil and gas industry was particularly hard hit by government regulations and taxes, and taxes on exports often defeated other objectives of the government and enterprises. Perhaps the worst situation for a Western company is to lose its ability to operate as happened to the oil producers, Phibro Energy and Anglo–Suisse. Their multimillion dollar joint venture, White Nights, with Varieganneftgaz was declared invalid by the Moscow arbitration court, evidently because its Russian partner was unwilling or unable to intervene.[46] This case illustrates the necessity of having effective and reliable Russian partners who can exercise influence in government circles, the supply network, and other critical areas.

Traditionally, Russian managers succeeded by utilizing networks of personal contacts rather than formal, legal mechanisms, and this situation will undoubtedly continue, at least until a workable code of business laws is developed. A Washington builder used such tactics, and by fostering relationships with the local militia, government, and other power centers, was able to avoid criminality and other problems.[47] His experience should be of value to other Western business managers.

3 *Build on the human resource strengths of Russian enterprises.*

The Russian managers surveyed in 1992 considered a capable work force as a major strength of their enterprises. Although the caliber of the work force had deteriorated for many state–owned enterprises by 1995, some Western companies described earlier have utilized this resource in their operations, and some, like Gillette, have even begun manufacturing in Russia.

Russia has an abundance of highly educated and experienced professionals. For instance, there are 1.7 million scientific workers and 7 million engineers, more than double the number in the United States.[48] A former Pratt & Whitney executive, in assessing Russia, noted, "The country is brimming with talent and technology that can be picked up at bargain prices and used to start new businesses. The real jewels are in the research bureaus, in the vast pools of knowledge."[49] Another American company, Connolly International, Ltd., creatively tapped this talent by employing a group of technical specialists in St. Petersburg to interpret U.S. product specifications such as those of Underwriters Laboratories. The objective was to gain approval by the government standards agency, Gosstandard, for U.S. products to be imported into Russia.[50]

In spite of work force problems which have persisted for many Russian enterprises, well educated, energetic and ambitious Russians, generally under age 40, have been identified as the key to the country's competitiveness, and have been hired by numerous Western firms. Virtually all those hired by Lotus Development Russia, for example, have been under age 35 since, according to one Lotus manager, "It's easier for a person with less of the wrong skills to learn more of the right skills."[51] Other Western firms may also be able to attract the best of Russia's human resources.

4 *Support training at all levels.*

The development of employees was a major objective of the 1992 respondents, while the lack of talented managers was seen as a weakness. The failing situation of many state–owned and newly privatized enterprises in 1995 made it clear that these managers had yet to acquire a significant body of new skills in market–based systems. To do so would involve substantial training and adaptation, and additional time would be required before such managers could become effective contributors to their organizations in the new economic conditions.

Western firms can help in management development by providing educational opportunities to Russian managers with whom they do business or employ. For instance, Baird Corporation, a manufacturer of spectrometers, has sent Russian managers to company training programs in Holland and the United States.[52] Otis Elevator, with several joint ventures in Russia, established links with Russian training institutions to ensure that training provided was applicable to local conditions.[53] Another avenue would be to support participation in management training programs such as those offered by Russian and Western business schools as well as international agencies like the World Bank's Trainers' Program on Enterprise Restructuring and Privatization.[54] Moreover, Western managers must themselves learn more about Russian business and culture, perhaps by attending programs such as those at Moscow State University, where Japanese managers have participated for the past several years.[55]

5 *Help Russian managers to understand marketing better, and to develop and market competitive products and services.*

Russian managers reported in 1992 that to succeed required good quality products or services, as well as appropriate technology and competitive pricing. Good marketing, however, was not seen as either a problem or a strength, nor as being important to business success. Failing to see customer issues and competition as problems, however, suggests a lack of understanding of what marketing is. This failure, based perhaps on their former monopoly positions, likely contributed to the deteriorated positions of enterprises as they entered 1995. An increased understanding of marketing would be essential to compete against better quality imported products and services, many of which owed their success to higher levels of technology.

Russian managers responding to the 1992 survey saw the technological levels of their enterprises as a relatively serious problem. Their lack of investment in R&D and new plant and equipment, however, prevented access to more advanced technologies, as well as difficulty in implementing technological change. This situation had only worsened by 1995.

Although some leading Russian advertising agencies have become considerably more sophisticated in the past three years,[56] it is still crucial for Western managers doing business with Russian enterprises to provide marketing expertise, and technological support when necessary. Some Russian firms have been able to offer high quality products and services, and would require only marketing assistance. For instance, Krasnyi Proletarii's exhibit of its products at an industrial trade fair in Chicago was facilitated by an American business contact with marketing expertise. Other success stories have occurred in the aviation industry, where technologically advanced products have been marketed abroad with assistance from Western partners. Aviastar, for example, had adopted an aggressive approach to marketing due, in part, to an agreement with HeavyLift of the United Kingdom to operate Volga–Dnepr Cargo Airlines. The venture's goal was to become a worldwide carrier, with HeavyLift leading the marketing efforts, while helping Aviastar managers learn the techniques.[57] In this case, both parties benefited from each other's strengths.

Likewise, the Moscow Tool Factory, MIZ, in collaboration with its German joint venture partner, realized about 10 percent of its business from foreign contracts. Their joint venture, Reform, supplied modern equipment and helped MIZ negotiate sales contracts with distributors in several Western countries, including the United States. "MIZ makes very fine tools to whatever specifications we demand and they're cheaper than tools made in the U.S.," stated the sales manager of Barit Corporation, a U.S. distributor which supplied MIZ tools to more than 30 North American companies.[58] As these successful ventures illustrate, opportunities exist for Western firms to add the necessary marketing and technological expertise to polish prospective Russian gems.

6 *Develop reliable sources of supply.*

The Russian managers reported in 1992 that obtaining reliable sources of supply was their major problem. This is hardly surprising in light of chronic shortages, the lack of clear alternatives to formerly centralized allocations, and difficulties with infrastructure such as transportation and distribution. As privatization progressed, such problems led many enterprises to revert to dependence upon their former state suppliers, in some cases even offering them part ownership of their enterprises to secure supplier support.[59]

Obtaining supplies also created problems for Western companies doing business in Russia. McDonald's, for instance, took the unprecedented step of developing and managing its own source of supply for virtually all food products. As Russian managers learned over the years, vertical integration became a mandatory strategy. McDonald's built their own food processing facility in Moscow and operated their own farms.[60] Ben & Jerry's Homemade, Inc. made a similar decision in their ice cream plant in Karelia.[61] In doing so, these innovative North American companies serve as an example for other Western firms.

Another creative way that Western companies may ensure supplies is to work with Russian enterprises that have a history of relatively reliable supply sources. Mikromashina of Moscow, for instance, experienced little difficulty obtaining plastic raw material and metal because it formerly operated within the well-supplied aviation industry. Mikromashina maintained its supplier network, and its U.S. joint venture partner, Nypro, Inc., also benefited as a result. Partnerships with reliable Russian suppliers can produce advantages for Western companies, who, in turn, can provide some stability for their Russian partners.

7 *Utilize creative financing methods.*

The 1992 survey respondents clearly understood the new importance of financing their businesses, noting that profitability had become necessary for success. They also ranked financing their businesses as a major problem. The difficulty with financing enterprises, whether state–owned or partially privatized, had several origins. The banking system was being radically restructured, the ruble remained highly unstable and became precipitously devalued, inflation was rampant, interest rates exorbitant, and stock exchanges and venture capital organizations were in a very early stage of development. In such a volatile financial environment, obtaining capital for investment was exceedingly difficult, and often impossible. Uncertainty about government subsidies, as well as tight credit, were other factors contributing to financial deterioration. These managers recognized that investment funds were virtually nonexistent, and the situation had deteriorated further by 1995.

Western firms must monitor closely which companies are likely to continue receiving government funds, and Westerners can utilize creative financial approaches to counteract the financial weaknesses of Russian customers and partners. Wang Laboratories Inc., for instance, was able to obtain payment in hard currency for computer hardware and software before the ruble became essentially convertible in 1994. The products were purchased by the Moscow City Council in 1993 through the Bangladesh–USSR trade agreement. Bangladesh had previously purchased six thermal power stations from Russia on a commercial credit basis, and it was agreed that an equivalent amount of goods would be exchanged in payment. Bangladesh firms had permission to negotiate contracts with Russian customers such as Moscow City Council, and through this mechanism, Wang was able to provide computers to a Bangladesh company which transferred them to Moscow. The Bangladesh firm received $2.5 million as their share of the original $25 million letter of credit, and provided Wang its $1.9 million.[62]

Such arrangements are circuitous, but can funnel hard currency payments to Western firms that would not otherwise be possible. A more direct method was used by Caterpillar Inc. which sold 295 heavy bulldozers to the Russians for $86.2 million in 1993. The payment from Russia's Gazprom was made possible by a direct credit from the U.S. Export–Import Bank.[63] Hard currency payment continues to be desirable in light of continued inflation in Russia and the serious instability of the ruble, despite ruble convertibility.

In summary, Western companies have to depend upon their own financial resources and creativity to conduct serious business in Russia since their partners will often be unable to contribute financially to business ventures. Some potential gems like Gazprom and LukOil, however, do possess the resources to create profitable opportunities for Western partners, and should be good investments. As obvious as it may seem, however, Western managers must be certain of who really controls the Russian enterprises they deal with. An ambiguous legal framework, coupled with the unfamiliarity of most Russians with private ownership, have produced situations where determining the true owners is exceedingly difficult. Many companies, similar to the traditional Matrioshka doll, are nested within one another, and finding the controlling party can be a puzzle.[64] Here, too, having reliable and capable Russian partners is crucial.

Conclusion

Western managers can play a profitable role in working with their Russian counterparts as privatization continues through the 1990s, provided they clearly understand the situation faced by Russian managers. "There have been bumps in the road, and there will be more in the future, but so be it," according to the chairman of Dresser Industries, a firm that has invested heavily in Russia.[65]

In spite of such "bumps," the Russian government once again confirmed its commitment to privatization in January 1995. The head of the privatization effort was replaced immediately after he called for nationalization of some key industries, a statement which caused concern among foreign investors.[66] This action followed the dismissal of the Central Bank's head, Viktor Gerashchenko, whose liberal credit policies and subsidies precipitated October 1994's "Black Tuesday" ruble crisis. His replacement, Tatiana Paramonova, is known for her belief in tight monetary policies. Her appointment was accompanied by the establishment of a new credit commission chaired by Anatolii Chubais, which is authorized to curb credits and public expenditures. Other developments in mid-1995 were the stabilization of unemployment reported to be 7.5%, an apparent stabilization of production and gross domestic product, a marked deficit reduction of nearly 10% of GDP in 1994, coupled with a strict 1995 budget, and a prediction by some experts of a drastically lower inflation level of 20% to 30% per year.[67]

In summary, some experts have seen many reasons to believe that the worst is over in Russia and that both the political and economic systems are stabilizing. Yet, if Russian enterprises are to make a successful transition to private status in a market-oriented economy, their managers, who have been essentially "muddling through" with their strategies,[68] will have to become more experienced and effective in formulating focused competitive strategies to take advantage of the apparently improving economic and political environment.

To accomplish this during the second phase of privatization, Russian managers will undoubtedly seek more opportunities for collaboration with Western businesses, and to attract from them sorely needed investment funds. "With care, most problems can be solved and profitable ventures created—provided an investor allows extra time and effort before expecting serious returns," advised a vice–president of McHugh Construction Company, a Chicago firm with four years' experience in Russia.[69]

The recommendations offered here suggest ways in which Western managers can work with their Russian counterparts to assist them, while at the same time, taking advantage of opportunities in that rapidly changing country. Privatization has created both pitfalls and potential profits, but if the rust is scraped away, some diamonds can eventually shine through.

The authors thank Drs. Alexander I. Naumov and Oleg S. Vikhanski of Moscow State University School of Business Administration, and Dr. Stanislav V. Shekshnia of Otis Elevator International, Paris, for their invaluable assistance in this research. "Diamonds and Rust" is the title of a song by Joan Baez.

Notes

1 See the Spring 1993 special issue (Volume 28, No. 1) of *Columbia Journal of World Business* on privatization initiatives around the world, especially Scott Thomas, "The Politics and Economics of Privatization in Central and Eastern Europe," 163–178. Refer also to A.F. Alkhafaji, "Privatization: An Overview," *Journal of Organizational Change Management* 6, No. 3 (1993): 24–42; "Survey—Business in Eastern Europe: Owners Are the Only Answer," *The Economist*, September 21, 1991, 10–14; "Eastern Europe Hesitates," *The Economist*, May 16, 1992, 13, 14; Ira W. Lieberman and John Nellis (eds.), *Russia: Creating Private Enterprise and Efficient Markets* (Washington, D.C.: World Bank), 1994.

2 Detelin Elenkov, "The Russian Aerospace Industry: Survey With Implications for American Firms in the Global Marketplace." *Journal of International Marketing*, forthcoming.

3 John Thornhill, "Foreign Speculators Raise the Stakes." *Financial Times*, October 10, 1994, 15.

4 Krystyna Krzyzak, "Milestones on the Capitalist Road." *Euromoney*, July 1994, 34–38.

5 Peter Galuszka, Stanley Reed, Stewart Toy, and Heidi Dawley, "A Russian Gas Giant May Get Some Foreign Fuel." *Business Week*, November 28, 1994, 70–71.

6 "The Eighth Sister," *The Economist*, October 15, 1994, 93.

7 Peter Galuszka, Stanley Reed, Stewart Toy, and Heidi Dawley, "A Russian Gas Giant May Get Some Foreign Fuel." *Business Week*, November 28, 1994, 70–71.

8 John Lloyd and John Thornhill, "Investors Set to Pour Into Russia if Rules are Tightened," *Financial Times*, October 10, 1994, 1.

9 Associated Press, "Yeltsin Pushes Plan to Revive Economy." *Boston Globe*, November 27, 1994, 16.

10 Chrystia Freeland, "Russia Set to Impose Tough Curbs on Spending," *Financial Times*, October 21, 1994, 1, 18.

11 *Economic Newsletter* Russian Research Center, Harvard University, June 20, 1994, 4.

12 Patricia Kranz, "Russia's State Sell–Off: 'It's Sink–or–Swim Time,'" *Business Week*, July 4, 1994, 46–47.

13 Jeffrey Sachs, "What Have We Learned About Rule of Law and Economic Reform in Russia?" Seminar, Russian Research Center, Harvard University, May 9, 1995.

14 "Not the Real Thing Yet," *The Economist*, March 12, 1994, 58.

15 Peter Galuszka, "'It's Like Climbing the Himalayas,'" *Business Week*, May 2, 1994, 106–107.

16 "Not the Real Thing Yet," *The Economist*, 58.

17 Dr. Alexander Naumov, deputy director of Moscow State University School of Business Administration. Personal communication, Boston, December 20, 1994.

18 Clinton O. Longenecker and Serguei Popovski, "Managerial Trials of Privatization: Retooling Russian Managers," *Business Horizons*, (November–December 1994): 35–43; Avraham Shama, "Management Under Fire: The Transformation of Managers in the Soviet Union and Eastern Europe," *Academy of Management Executive*, (February 1993): 22–35.

19 *Economic Newsletter* Russian Research Center, Harvard University, August 5, 1994, 4.

20 Ibid.

21 John Thornhill, "Russia Says Economy Will Overcome Instability," *Financial Times*, October 25, 1994, 2.

22 John Lloyd, "Twisting and Turning." *Financial Times*, November 8, 1994, 16.

23 Chrystia Freeland, "Russia Set to Impose Tough Curbs on Spending," *Financial Times*, October 21, 1994, 1, 18.

24 "Conversed Enterprises: Economic Situation and ospects for Development in 1994," *Business World Weekly*, (no. 35/128): 10–11.

25 "The Eighth Sister," *The Economist* October 15, 1994, 93.

26 Nicholas Denton and John Thornhill, "Russian Gas Giant Approaches Investors," *Financial Times*, December 5, 1994, 15.

27 Peter Galuszka, "'It's Like Climbing the Himalayas.'" *Business Week*, May 2, 1994, 106–107.

28 *Economic Newsletter* Russian Research Center, Harvard University, July 12, 1994, 2.

29 *Economic Newsletter* Russian Research Center, Harvard University, December 21, 1994, 1.

30 The analysis for 1995 is not based on a survey of managers as in 1992, but on published reports and interviews we conducted with Russian managers and academics and other Russian and American experts in the fall of 1994. Neither analysis purports strict statistical validity, but both can be considered reasonable assessments of conditions in Russian enterprises and their environment. For additional information on the 1992 survey see Daniel J. McCarthy and Sheila M. Puffer, "Russian State-Owned Enterprise Managers View the Transition to Private Enterprise." *Journal of Private Enterprise* 10, no. 1 (1994): 92–107.

31 John Lloyd, "Storms Break Russia's Long Political Calm." *Financial Times* October 21, 1994, 2.

32 *Economic Newsletter* Russian Research Center, Harvard University, December 21, 1994, 1.

33 Simon Clarke, Peter Fairbrother, Vadim Borisov, and Petr Bizyukov, "The Privatisation of Industrial Enterprises in Russia: Four Case-Studies." *Europe-Asia Studies* 46, no. 2 (1994): 179–214.

34 Ibid.

35 Richard M. Hammer, Hans H. Hinterhuber and Justus Lorentz, "Privatization – A Cure For All Ills?" *Long Range Planning* 22, no. 6 (1989): 19–28.

36 Daniel J. McCarthy, Sheila M. Puffer and Peter J. Simmonds, "Riding the Russian Roller Coaster: U.S. Firms' Recent Experience and Future Plans in the Former USSR." *California Management Review* 36, no. 1 (1993): 99–115; Peter Galuszka, Patricia Kranz, and Stanley Reed, "Russia's New Capitalism." *Business Week*, October 10, 1994, 68–80 (*passim*).

37 Paul Hofheinz, "Rising in Russia." *Fortune*, January 24, 1994, 92–97.

38 Yuri Petchenkine, "New Boost to Privatization in Russia." *East/West Executive Guide* 2, no. 7 (July 1992): 19–22.

39 Peter Galuszka, and Patricia Kranz, "Crunch Time in Russia." *Business Week*, February 7, 1994, 42–43.

40 *Economic Newsletter* Russian Research Center, Harvard University, June 20, 1994, 2.

41 "Distiller of Stolichnaya Is Insolvent, But Russia Wants to Avoid Liquidation." *The Wall Street Journal* August 17, 1994, A9.

42 David Brooks, "You're Privatized. Now What?" *The Wall Street Journal*, April 23, 1993, 4.

43 Ibid.

44 Simon Clarke, Peter Fairbrother, Vadim Borisov, and Petr Bizyukov, "The Privatisation of Industrial Enterprises in Russia: Four Case-Studies." *Europe-Asia Studies* 46, no. 2 (1994): 179–214.

45 Steven Erlanger, "To Russia, With Good Ol' American Know-How." *The New York Times*, June 13, 1993.

46 *Economic Newsletter* Russian Research Center, Harvard University, July 12, 1994, 3.

47 Robert Keatley, "Russia's Tales of Horror May be Hiding Golden Opportunity, Some Investors Say." *The Wall Street Journal*, June 17, 1994, A9.

48 M. Kublin, and J. Shepherd, "Technology Transfer from the Commonwealth of Independent States to the United States." *Industrial Management* 35, no. 2 (1993): 26–29.

49 Paul Hofheinz, "Rising in Russia." *Fortune*, January 24, 1994, 92–97.

50 John Baranofsky, Connolly International, Inc. Personal communication, Boston, February 3, 1994.

51 Patricia Kranz, "The Young and the Ambitious." *Business Week/21st Century Capitalism*, November 18, 1994, 128.

52 Sheila M. Puffer and Jack Medzorian, "Manufacturing Spectrometers in St. Petersburg." In Sheila M. Puffer (ed.), *Russian Management Revolution: Preparing Managers for the Market Economy* (Armonk, NY: M.E. Sharpe, 1992), 248–257.

53 Stas Shekshnia, "Otis HR Finds Challenging Times in Eastern Europe." *Human Resources Newsletter* United Technologies Corporation, 2, no. 2, (Fall 1993): 1.

54 *Trainers' Program on Enterprise Restructuring and Privatization* (Washington, D.C.: Economic Development Institute of the World Bank in Collaboration With the International Business School, Moscow, April 12–23, 1993.)

55 Jim Vail, "Russians Learn the Art of Japanese-American Business." *The Moscow Tribune*, October 12, 1994, 7.

56 David Filipov, "Russia Seeing a Bit of Madison Ave. on Gorky Street." *Boston Globe*, December 26, 1994, 63.

57 Detelin Elenkov, "Russian TNCs in Global Competition: Their Origin, Competitive Strengths and Forms of Transnational Expansion." *Columbia Journal of World Business* 30 (1995, forthcoming).

58 Fred Kaplan, "Russia's 'Red Barons.'" *Boston Sunday Globe*, February 20, 1994, A1–A2.

59 Simon Clarke, Peter Fairbrother, Vadim Borisov, and Petr Bizyukov, "The Privatisation of Industrial Enterprises in Russia: Four Case-Studies." *Europe-Asia Studies* 46, no.2 (1994): 179–214.

60 Oleg Vikhanski and Sheila Puffer, "Management education and employee training at Moscow McDonald's." *European Management Journal* 11, no. 1 (1993): 102–107.

61 Iris Berdrow and Henry W. Lane, "*Iceverks (A): Ben & Jerry's in Russia, and Iceverks (B): Summer of 1993.*" (London, Canada: University of Western Ontario Case and Publication Services 1995).

62 "Wang Translates Russia–Bangladesh Trade Agreement into Real Money." *East/West Business and Finance Alert* (Concord, MA: WorldTrade Executive, Inc., July 1993) 4.

63 "Tenders and Investment Opportunities." *East/West Business and Finance Alert* (Concord, MA: WorldTrade Executive, Inc., July 1993) 6.

64 David Brooks, "You're Privatized. Now What?" *The Wall Street Journal*, April 23, 1993, 4.

65 Steve Fainaru, "Russia's Risk Outpaces Opportunity." *The Boston Sunday Globe*, January 23, 1994, 75, 81.

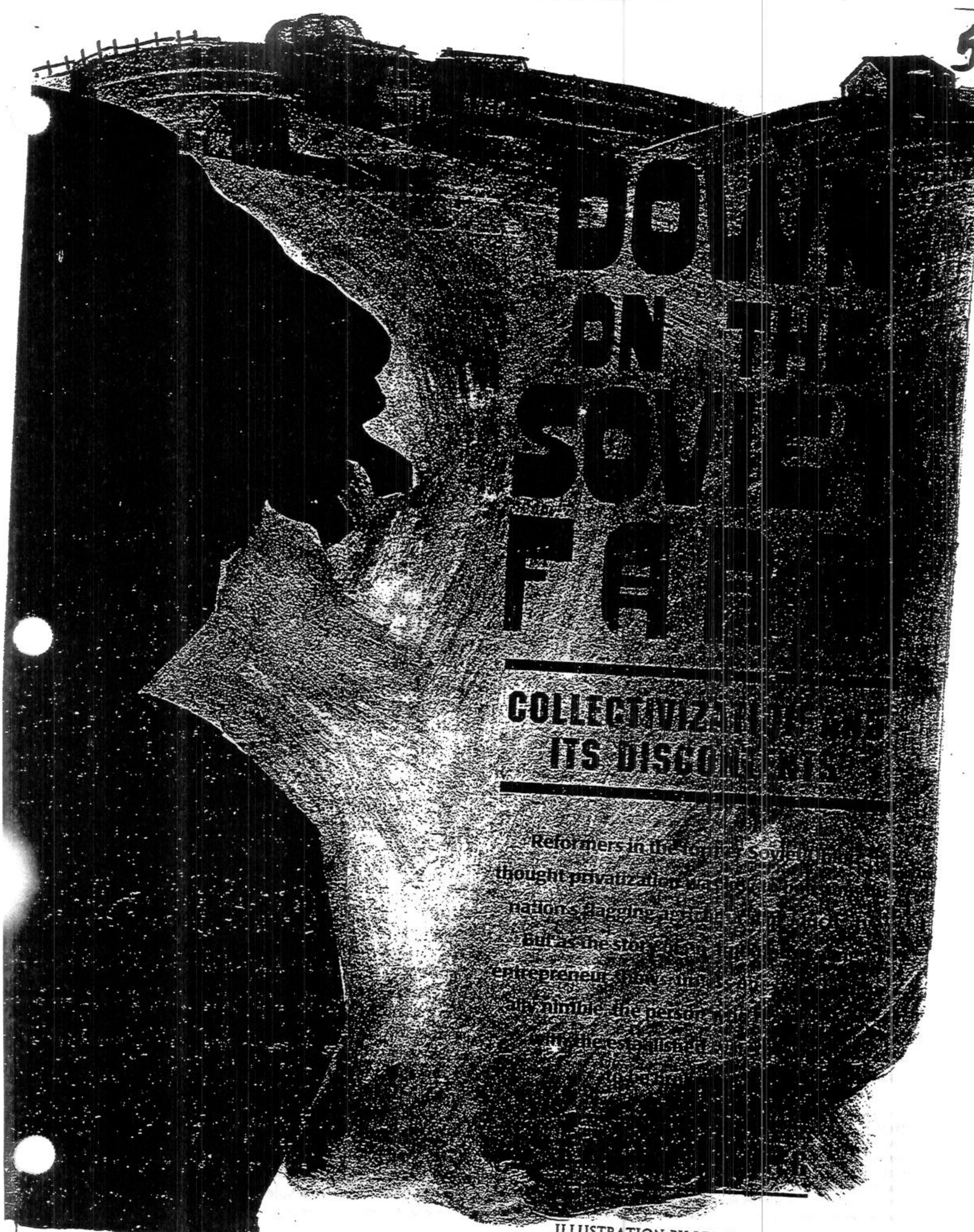
66 "Privatization Chief in Russia Ousted." *The Boston Globe*, January 25, 1995.

67 Anders Aslund, "Ruble Awakening: Why the Economic News From Russia is—Surprise!—Good." *Washington Post*, April 23, 1995; Jeffrey Sachs, "What Have We Learned About Rule of Law and Economic Reform in Russia?" Seminar, Russian Research Center, Harvard University, May 9, 1995; Robert Watters, "Practical Solutions for Stimulating Russia's Private Sector." In Reiner Weichhardt (ed.), *Privatization in NACC Countries* (Brussels: NATO, 1994) 275–279.

68 Charles Lindblom, "The Science of 'Muddling Through.'" *Public Administration Review* 19 (Spring 1959): 79–88.

69 Robert Keatley, "Russia's Tales of Horror May be Hiding Golden Opportunity, Some Investors Say." *The Wall Street Journal*, June 17, 1994, A9.

53



DOWN ON THE SOVIET FARM

COLLECTIVIZATION AND ITS DISCONTENTS

Reformers in the former Soviet [illegible] thought privatization [illegible] nation's flagging a[illegible] But as the story [illegible] entrepreneur [illegible] ally nimble, the person [illegible] with the establishment [illegible]

ILLUSTRATION BY LES KANTUREK

TECHNOLOGY REVIEW APRIL 1996

WHARTON REPROGRAPHICS

I ASSURE you," said Lev Abramovich, ancient official greeter, keeper-in-line of foreign journalists, and apologist for Ivanovo Region agriculture, "I travel around, and the barn you are about to see at Collective Farm Trudovik is the worst in the area." My official Communist keeper for this leg of my journey across the Soviet Union in 1989 didn't approve of what his bosses at *Gosagroprom*—the monolithic, all-controlling, late-Soviet superministry of agriculture—had assigned him to show me: the efforts of one of the nation's first halfway independent farmers in decades.

We piled in an old jeep, and Nikolai Petrachkov, chairman of Farm Trudovik, drove us across one of the bumpy wheat fields in his oversized dominion. I had been encouraged to visit Nikolai by Viktor Lishchenko, a key reform Communist government adviser I had met a year earlier on a prior reporting trip. Inspired by the government's new policy of *glasnost*, he supported my proposal to write about Soviet collective farming in a time of presumed change and examine why a nation that had developed a nuclear arsenal and whose farms stretched from Norway to Korea across 11 time zones suffered nearly empty food shops and imported half the wheat in every loaf of bread.

Nikolai stopped the jeep and spoke to several men who lounged beside a halted combine—a standard work brigade on a standard day. This was expansive, flat country, near the town of Shuya, Ivanovo region, some 200 miles northeast of Moscow, and a world apart. Even in midsummer the air felt thin, a breeze away from the Arctic. Far across the big grain fields, a ribbon of low pine woods underlined the broad gray sky. The vista reminded me of Maine's potato counties.

We drove on. Nikolai spoke, through an interpreter, about the future of Soviet farming. Unlike the typical collective farm chairman (to this day) who would preserve a centralized system that had long provided a comfortable niche, Nikolai envisioned far greater autonomy for individual farmers. He was eager to show the world the early results of his experiment in leasing one of Farm Trudovik's dairy barns to a brigade headed by a bright and ambitious young worker named Volodya Kurikhin.

We pulled up in front of Volodya's tumbledown cattle shed. In fact, we'd passed many barns in worse condition on the 45-minute drive out to Shuya. Lev was not above spin control. "The barn sags," he said. It did, but the roof looked sound. I was naive enough to feel excited. Nikolai had praised Volodya so highly I felt as though I were about to meet the hope of Russia.

Volodya wasn't prepossessing, which was proper for a savior.

Volodya Kurikhin

Nikolai Petrachkov

His gestures were restrained—a shy glance at the visitor, a hand barely proffered. Compared with the charismatic and routinely hearty Party men, the ideological guardians of collective agriculture with whom I'd had to keep regular company during many months of travel across the Soviet Union, Volodya seemed a faint hope. He stood just under six feet tall, with short, sandy hair and a completely cockeyed grin—a right-sided smile that he delivered with head listing to port.

How long, I asked, had he been a member of the Farm Trudovik? He shook the listing head. "I'm not a member. On my residence papers, I am still registered as an electrician from Shuya. I worked on many new barns, including the one that sits over there," he pointed at another building. "I finished and moved on to other farm wiring jobs. But in May I came back here to milk cows."

We strolled across his barnyard and at my suggestion clambered up 10 tiers of hay bales to the tin roof of the big, open shed. It was clean and quiet. I arranged a bale as a desk, opened my laptop, and wrote looking out across Volodya's fields. It was a splendid, sunny afternoon. We had a good view. But the main reason Volodya and I and

---

MARK KRAMER is the author of Travels with a Hungry Bear *(published in April by Houghton Mifflin), from which this article is adapted. His previous work includes* Three Farms: Making Milk, Meat, and Money from the American Soil *(Atlantic Monthly Press, 1980) and* Invasive Procedures: A Year in the World of Two Surgeons *(Harper and Row, 1982). He teaches a graduate seminar on literary journalism at Boston University, where he is writer-in-residence and professor of journalism and American studies.*

PHOTOS: MARK KRAMER

an interpreter had climbed the hay mow was to get clear of Lev, who, in his seventies, didn't climb bales. He hung out down below, though, where he might overhear the conversation. We spoke quietly.

"I spent five years as an independent builder," Volodya said. "I hated working eight to five, with a time card. I'm like a cossack—independent." He was in the small first wave who'd decided to answer Gorbachev's ambiguous call for some independent farmers, he said. "I didn't know much about how to farm. Television shows some Western farms for a minute now and then. I've picked up a little from that.

"The problem," Volodya said, "is that everything—even the cows—still belongs completely to the collective farm. The farm can take them back and controls all decisions about them. We can't cull out a single bad cow and send her to be slaughtered without the approval of the collective-farm zootechnician. We have a cow as big as a railroad car who gives hardly any milk. She just eats. If I had my way, I'd sell her for meat and buy a calf of good breeding with the money. We wrestled back and forth for months before the zootechnician let us get rid of 10 terrible cows. There are 10 more we would sell if we could. But the zootechnician here is just looking for a big cow population so she can write down the census number and fulfill her quota; she's not looking for quality.

Unlike the typical collective-farm chairman, who would preserve a centralized system that provided a comfortable niche, Nikolai Petrachkov envisioned far greater autonomy for individual farmers.

"I also lack high-protein cow feed. That would make my cows give more milk. But the chief economist doesn't care about growing quality feed; she's just got crop quotas to count. After just a few months, I have proved that a higher yield is possible. In this old barn, we're already producing more milk than the regular dairy brigades provide here on Farm Trudovik. But it will take us three years to make a good herd—if they let us. I'm pushing the collective farm to double my land. With our number of cows, the norm is about 250 acres, and we don't even have half that.

Barn at Farm Trudovik

Lev Abramovich

"I need the chance to grow my own feed. I want to grow other types of grain than what they grow here—and have several fields in a rotation. I want to grow turnips, and perennial grass, and vetch. If I grew oats, which are in short supply, I could sell them and maybe use the money to buy other feed.

"There also isn't enough technical information for a small operator like me. We need foreign sources on that. We have lots of Russian handbooks about how to farm, but still the situation doesn't improve. It deteriorates."

In the midst of Ivanovo region's agricultural mayhem, Volodya Kurikhin burned with a fervent ambition unusual in the USSR. He was near to winning the rarest privilege in the nation: a chance to do a complex job very well. We climbed down from the haymow, and there was Lev right below us, hustling away lest we imagined he'd been trying to listen in.

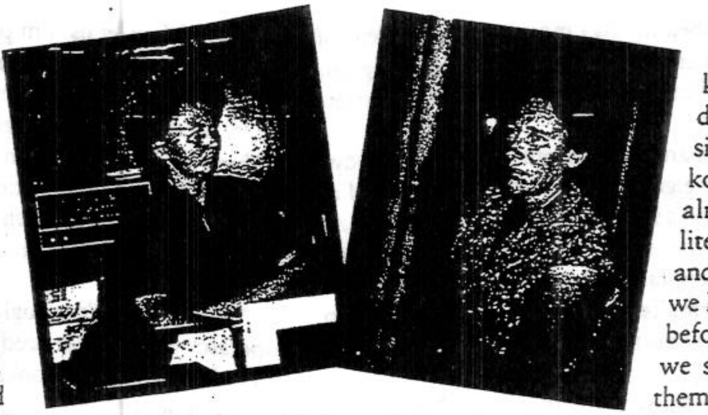
Volodya had organized a tiny office for himself inside his rundown barn. Cows lowed and munched cud outside the doorway. He'd slept there the night before—a greasy old quilt and crumpled newspapers spotted with food lay on a cot. Farm records were strewn on a board table. He was attentive to his job. I told him he was the first ambitious guy I'd met in weeks. He smiled grimly and said, "Some of my crew feel as I do, although one or two think this work is just a job, same as everywhere else."

His next line could have been scripted for a U.S. Information Agency propaganda film: "But even if someone is interested only in money, he'll still work hard for us. Because he knows that if he does, he'll get much more money. Because I even sleep here at night next to the cows, I'm in a good position to say to others on the crew, 'See here, I work hard and so should you.' " Volodya sounded like a manager.

He also sounded weary. His neighbors on the farm, working for their usual uninspiring wages, had seen Volodya's enterprise as a private giveaway of state property—their property. "Nikolai Petrachkov, the chairman of Farm Trudovik, has provided us with good land, a tractor, other equipment," Volodya said. "At the beginning I felt he was the only one who supported us here. Now I feel the situation is becoming more normal. Other people in the community who were envious are getting used to us. Their tempers have subsided. They are paying less attention. But I'm still doubtful that politics will change in the whole country, or even here on the farm."

Hoping to protect their novel arrangement against such politics, Nikolai and Volodya drew up a contract for the use of this barn, a kind of folk-lawyerly attempt to set down the rights and duties of both lessor and lessee, and they filed it in the office of the quota-conscious farm economist. The contract essentially created a mutually imposed zone of law in a nation with little business law. Contracts with workers were rare and, when they existed, more often amounted to no more than a sentence or two asserting—that workers would work and be paid.

This mongrel, homemade, topsy-turvy manifestation of both sides' wish for law was Nikolai's attempt to back himself, to render more legitimate and safer his authorization of an ideologically iffy and locally unpopular enterprise. He'd courageously gone ahead and supported Volodya in a time of new *perestroika* possibilities and continuing uncertainty. And he had done this in a place where private effort for private gain was considered selfish and immoral. Such doctrine was drummed into children all through school.



Farm economist Liubova B. Zotina (left) and chief zootechnician Nadezhda A. Shashkova

The contract, redolent of the time and place, was typed sloppily on a half-dozen sheets of yellowing paper. It specified assorted obligations of the lessee, Volodya, which included paying the rent, producing a quota of milk, inseminating and caring for the cows, maintaining the equipment, and removing dung. The duties of the lessor, Nikolai, included buying the milk and providing the barn, barn repair, land, well, milking equipment, and veterinary service.

Dated 10 May 1989, the contract would be valid for five years. Although it stated that "after every year, the treaty should be looked into again and some changes can be made," it stipulated that both sides must agree to any changes. It also provided that "if we have a disagreement over the execution of the agreement, we can bring it to the attention of the Shuya Court."

"Volodya is a good farmer now," Nikolai Petrachkov said. We sat at the supper table with his wife, Taisa Stepanovna, who was chief economist at Farm Gorky, the next farm down the road. We worked through a new delivery of special-guest food that had been delivered by Party officials that afternoon for the comfort of the American Delegation (me). "He was not one when he started. After helping me wire a new dairy barn, he proposed a small farm, in line with Gorbachev's suggestion that we establish lease brigades. After many meetings, I felt he was serious, so we worked out the contract you saw. We gave him and his crew the old, abandoned barn. Our zootechnician selected 60 animals—frankly, they were the worst on the farm. But that's what she gave them.

"This summer," said Nikolai, "Volodya and his brigade are working 107 acres with two tractors we also gave—no, rented is the word. They pay rent to Farm Trudovik. The farm economist knew how to figure out the rent for the tractors and barn. This is not completely new. As for renting land, that is confusing. In the end, the lease-brigade members earn about twice what others at Trudovik earn, but work longer, harder hours. They produce more.

"We buy all their milk, for an average of 35 kopeks a liter. The state pays us 38 kopeks for each liter we deliver. We get another subsidy that makes our total 54 kopeks, so we actually profit almost a ruble for every 5 liters we buy from Volodya and sell to the state plant. And we have fewer headaches than before. They perform so well, we seldom hear from or see them. They take care of growing and finding their own feed. They've already improved the building, which was poor. They've got the tractors running. They have almost never shown up at my office with unexpected problems.

"On our books, we still pay out exactly the salary that we did before for maintenance of those 60 cows," Nikolai said. "But we pay 6 people instead of 12. And I have to say, that makes some of the people in our town angry at them, and at me. 'Look,' the people say, 'the land belongs to the collective farm. The cattle also belong to the farm. So why do you allow those people to make a killing out of our common property? They don't share their money with everybody. How can you allow that?' But not everybody complains, and it's not hard for me to handle. This experiment should be here."

Taisa Stepanovna was more direct. "The complaints come from loafers, too lazy to work hard," she said. "They're a kind of union of sluggards, young and old, men and women, paid all their lives for the literal execution of their orders, without a care in the world about quality."

"It bothers people who earn half as much," said Nikolai, "and don't see Volodya in the barn 14 hours a day, then sleeping right there, getting up during the night to tend new calves. Envy has no logic. These others could work harder and earn more, too. But Volodya and his crew are the only ones who have asked for a lease brigade. I hope more people will see that this way is good for our farm and country. In fact, there are huge sections of abandoned land that should be given to small farmers. But as long as we don't have basic laws on land or property, I am not even sure that what I have done is legal."

After breakfast, in the middle of my week's visit, I sat in Nikolai's office at Farm Trudovik. Nikolai read mail, responded to questions from crew chiefs and aides who ran in and out, and planned out his day of keeping the harvest going and construction projects working. Eventually he found time to chat. "Volodya's farm has been in operation just three months," he said. "Let's review its production numbers." He picked up the phone and gave a few commands.

The farm economist strode in. Liubova B. Zotina sat with a straight back and spoke with clipped, precise speech: "To date on the special unit, milk per cow: consistently 12 percent overfulfillment of their plan. In May, June, and July, projected average yield per cow was 325 liters. Actual average was 364 liters." What she did not say—although Nikolai looked up the numbers later—was that Volodya's group also outstripped Farm Trudovik's production. In May, June, and July, Trudovik cows averaged 284 liters of milk. Given the worst cows and the worst building, Volodya's crew started out doing 28 percent better than the collective work brigades. Increases in yield on dairy farms are hard won. The numbers were impressive.

Nikolai called in the chief zootechnician, Nadezhda A. Shashkova. She sat, ankles crossed and face stern, and reluctantly recited her digest of Volodya-related facts: "He's got basically the same quality herd that we have on the rest of the farm. On my authority we have sent for butchering 10 of 60 cows, and I am looking at eliminating another 6. Our average cow lasts six or seven years. By regulation, we may cull only barren cows or those with atrophy of the udder. Otherwise, we are required by law to keep every one and push milk production even in less-productive animals." This amazing regulation made no sense for any farm wanting to maximize milk production. American cows last about four years in herds. Such turnover is a part of high production.

Just three months after being given the worst cows and the worst building, Volodya's lease brigade was 28 percent more productive than the collective crews at Farm Trudovik.

So I asked the zootechnician a question she may have found bold, or even rude: how did she feel about doing her job less well than she knew how, in order to make the record look good? Surprisingly, she smiled. "I feel frustrated and terrible after many years of work in our agriculture. We've always had to follow certain procedures contrary to science and common sense. In the past few years, under Gorbachev, we've started to breathe a little more freely."

I asked if she was surprised that Volodya had gotten good results so quickly. But Nikolai, a kind man, saved her additional embarrassment by speaking up: "I hope it's the first bird of summer. I can't encourage others yet, with no law to back me up. When there is, I'll start more lease brigades. We can't raise productivity following the old path. Lease brigades seem like the surest available way toward change. Look at those statistics: In labor productivity, these city boys already do as well as our most experienced milkmaids! They even harvested and stacked their own hay, I see, which somehow isn't part of their productivity data—and which milkmaids don't do."

Aleksander Sorokin

"I'll add hay activity," the economist said glumly. She saw me studying her and offered a slight, strained grimace. The farm zootechnician's face mirrored the expression. The two were disapproving underlings. They didn't like Nikolai Petrachkov's sponsorship of Volodya, and were clearly edgy in this candid discussion.

Their overt opposition impressed and puzzled me. I later came to understand that their attitudes were the norm. Nikolai was the adventurer. His outspoken support of the idea of efficiency, of more pay for more work, of maximizing asset management, jeopardized the communal sense of equality and stimulated resentment of Volodya's private advantage. As little as communal equality represented the actualities of Soviet life, it was operant patriotic myth.

The sourness of the economist and zootechnician was in the air every time I ran across them during my week on Farm Trudovik. I was glad to leave, and during my first few days back in America, I found myself, with the fervor of a Jay Cee, drinking in every glimpse of good work well rewarded.

The next summer, I was back in the USSR and had a few days for a hurried trip to Ivanovo. I wanted to see what a year and a half on the job had brought to Volodya's lease brigade. I took the train to Ivanovo with an interpreter named Yelena, a graduate student suggested by Viktor Lishchenko, the senior adviser on agriculture to the reform Communist government who had helped plan some of my Soviet travels.

We got off the train and went to see Aleksander Sorokin, the regional head of agriculture with whom I'd drunk vodka and had a heartfelt chat one evening during my previous visit. "How's Volodya?" I asked. Sorokin answered, "I've heard he's got family troubles. His wife—I heard his wife complained about his farming."

"But I have an interview on tape," I said. "He says how proud she is and that she approves of his hard work."

"Well, ask Nikolai Petrachkov," said Sorokin, shrugging, dismissing Volodya and bringing up instead a blue fox farm that was a personal fascination.

Yelena and I drove the 45 minutes out to Shuya. At high noon at a restaurant in town Nikolai greeted me like an old friend, clapping me on the back, grinning; I'd rarely felt so warmly welcomed anywhere. As on our first meeting, he looked like a country and western star. His pompadour yet waved, with no gray in it. At lunch, Nikolai and I discussed

America, not business. Then we jounced out of town in his jeep—but not to Farm Trudovik.

Instead we went to a ground-floor office in a cement administration building on Farm Gorky, the next farm down the road, where Nikolai's wife worked. Over the winter, the prior chairman had retired, and Nikolai had been invited to leave his previous post and join his wife in running it. I began to suspect what had happened to poor Volodya.

Road behind Farm Trudovik

We sat at the long table that nearly filled Nikolai's new, smaller office. A portrait of Lenin hung on the wall behind him. Nikolai, reluctantly at first, filled in events. Aleksander Sorokin had offered him the job: "Sorokin said it would be difficult for an outsider to take on."

Taisa came in and out, half listening, working at her own projects. "I considered staying on at Farm Trudovik," Nikolai said, "to finish the projects I'd started. But I couldn't resist the new job." Not only had he grown up on Gorky, but it was larger, more fertile, and richer, and he liked having his wife guarding his back as farm economist, in place of the oppositional staff at Trudovik. "I lacked complete harmony there," was his polite understatement. "But the return to my home farm brought one painful regret—the move let loose troubles that brought Volodya Kurikhin's farm down."

Volodya had needed Nikolai's protection right from the start—that was normal for a reform experiment amid the tangle of bureaucracy connecting every job. A Soviet rule: if a corner of the economy worked, it was a protected exception. Like the wealthy collective farms I'd visited whose chairmen had found patrons high in the Party, Volodya's lowly but prosperous lease brigade had worked out because it had special protection. It had started after Volodya impressed Nikolai, who'd impressed Sorokin, who'd impressed some big boss in Moscow, perhaps Viktor Lishchenko, forming a chain of personal impunity that warded off bitter functionaries and envious neighbors who hated communal resources producing private gain.

Nikolai's move over to Farm Gorky broke Volodya's chain of patronage. Farm Trudovik's new chairman was "a careerist—a conservative, normal chairman," Nikolai said, "a good chairman who knows how to farm and who was educated at a fine agricultural institute. He had previously managed another collective farm, and is younger, so he'll be able to build Farm Trudovik for years." Nikolai smiled.

Taisa, who had sat down to hear this part of the tale, stated the sad truth: "Nikolai's replacement did not believe in lease brigades." She was angry and bold. "The ladies you talked to with my husband in the office last year—that zootechnician, the economist—would have said the opposite of what they told you if they'd been alone. They'd have told you that Volodya's performance was mediocre, that allowing an outsider like Volodya on their land, especially to make such big money, was improper. The zootechnician had a big fight with her own husband before he himself joined that brigade.

"Nikolai couldn't throw his full support behind Volodya. He'd felt worried with no laws on property or land. Those laws still haven't come through. Nikolai felt the future of Soviet farming lay with the lease brigade—and private land, too. He had wanted to transfer Volodya to a better barn."

Nikolai spoke up: "Gorbachev recognizes that our main problem is that people can't take initiative. It is crucial. The case of Volodya demonstrated that things can change if you allow people to take initiative. I'd buy my collective farm if I had money and law permitted. It's a pity that 70 years of Soviet power haven't given the average Soviet an adequate philosophy so all collective farm members would really take responsibility." I'd often heard this sentiment expressed on my Soviet excursions, and it always made me sorrowful. I thought it an accurate lament about the default selfishness level of humankind, which capitalism plays to, but socialism exceeds.

"Now Nikolai doesn't dare create other lease brigades here" because Volodya's deal left normal work brigades seething, Taisa said. "Nikolai was frustrated that he couldn't even back Volodya strongly last year. There was enormous resistance to Volodya."

"Volodya came to me when his troubles over there were just starting," Nikolai recalled. "He asked me to let him lease an old barn on a back lot here at Gorky. We'd had problems with some workers on that farmstead, but we solved them, and I couldn't give it to him. I never imagined he would act drastically and suddenly drop farming. He'd done it so well. I told him to be patient, and things would be OK. Volodya did not stop by to see me again. I think his hopes were shattered."

We sat drinking tea, and Nikolai said abruptly, "I heard exactly what happened, and I'll tell you. The people who envied him just dismantled his farm. He earned too much. Volodya's crew always paid their lease fee precisely, in accordance with each detail of the leasing contract I showed you. They followed the agreement to the letter. But as soon as I left, the new chairman jacked up the land rent and added other fees.

"I'd promised to make further repairs on Volodya's barn at the farm's expense—it was so old, and was farm property, and he paid rent. The first thing they said to Volodya was, 'You must repair it now, at your own expense.' The new chairman piled on so many conditions that the lease brigade had to stop work and move away. It was a shame. They'd grown high-quality cow feed. They'd increased production. Well, now he's building houses in a cooperative construction brigade." Nikolai seemed caught by the waste of it all. He sat silent for a minute.

With Nikolai's special protection, Volodya's lowly lease brigade prospered. Without it, the reform experiment was quickly doomed amid a tangle of bureaucracy.

"Nikolai," I said, "let's drive down the road, find the new chairman of Trudovik and that farm economist and zootechnician, and ask for their account of what happened." Nikolai looked up, surprised. He thought for a moment. Then he grinned and hustled us into the jeep.

The flat, drought-browned wheat fields between the two farms blurred by in a few minutes. We jounced up to the dooryard of Farm Trudovik's office. Nikolai held up his hand. He would not go in. He would oblige the American Delegation in that time of maximal *glasnost* and he would enjoy imagining the conversation inside, but he would not witness the second-guessing of a hack chairman who'd been ensconced by higher-ups.

I opened the door to the office. The secretary wore a cotton housedress, white with little red and yellow flowers spaced evenly on it. "American journalist here to visit with the chairman," my translator, Yelena, announced. Then I was in his office. We had stumbled in at the right instant.

The chairman was compact and gray-haired, with a pleasant enough smile. He held out his hand and introduced himself: "Drozhin—Vladimir Anatolievich Drozhin." He seemed amiable and poised enough to be cordial, even though he must have been bewildered by this surprise visit. He introduced the economist and the zootechnician, whom I recognized from the year before when they'd worked for Nikolai, and who now sat at the new chairman's side. They recognized me, too. We shook hands.

This chairman kept two likenesses of Lenin in view, a portrait and a statue. The office must have already looked like this shortly after the farm started in the 1930s. It had a linoleum floor and walls thick with drab paint. I'd interrupted a staff meeting. The chairman dismissed all but the two women. I explained to him that I'd spent a week at the farm a year earlier and hoped to catch up on events. How was his new job going?

Drozhin: "It's unrealistic to steer a course before getting to know the people. People here seek order, but a manager can't simply command these days. We must find the instruments to make people do what they want to do."

This was a familiar Soviet voice, that of a standard bureaucrat. I asked about lease brigades.

Drozhin: "We try turning people toward leasing so they can earn more, and so they will have to remember where money goes. Leasing evokes the interest of workers in the final results..."

I felt cranky: "Where's Volodya Kurikhin this year? I spoke with him last year."

Drozhin: "He's not here anymore. He is in the town of Shuya."

"When I was here last year, he seemed so productive and determined."

Drozhin: "It was his old barn. It was uncomfortable for the calves in the winter. We had a warm winter, so he survived, but there were not proper conditions for milking there. I suggested a new barn at another sub-farm, with 200 cows. But he didn't agree."

"You offered a similar contract? Same deal?"

Drozhin: "Probably—some differences. They should pay for using pasture to graze cows, which they didn't do under the last contract. And the cost of land was not properly valued in their lease, so the lease was not properly valued."

"In other words, too much of the profit went to them?"

Drozhin: "Yes, almost all the profit went to them. The moral effect on the community was bad. Maybe my reason was not just economic."

The farm economist, Liubova Zotina, spoke up, offering another angle on Volodya's demise: "He left because he wanted private enterprise," she said. "He wanted to buy his own cattle, and not mix with the business of others. He can't even find a steady area of employment. He now works outside agriculture. Nobody knows where—something not very down-to-earth."

Nadezhda Shashkova, the farm zootechnician who had denied Volodya cow-culling rights, but whose own husband had joined the lease brigade, also chimed in: "My husband has returned to his place— in the collective's tractor-driver's brigade," she said. "It was impossible to live that way. Everybody envied him. There were rumors. People groused. Was I helping him to get special feed for the cows? I would not have let my husband keep working with this brigade, even if it hadn't halted."

Chairman Drozhin interrupted with his own further complaints: "They assumed no material responsibility for what they were doing. Leasing means we should get part of the profit of the tenant. We didn't get any. So we really had no payment for leasing, and if you do not have stable leasing payments, how can you have leasing at all? But you should understand that nobody drove them away. Nobody closed their farm."

The mysterious deity who opens doors for wanderers chose that moment to intervene: The chairman picked up his phone, spoke for a moment, and announced, "In a moment, a visitor will see you." And soon, in lumbered a tall, thick blond man in his thirties with a mild smile and serious, piercing blue eyes. "Loskutov," the man growled, "Aleksander Vladimirovich Loskutov."

"He was on the brigade with Volodya," said the chairman. "He lives here, and he again works on the farm in a regular brigade. You may talk to him. Ask him anything. We have nothing to hide."

Loskutov also recalled my visit to the barn the previous summer. "What happened to your lease brigade and to Volodya right after Nikolai Petrachkov left?" I asked. There wasn't much room for subtle questions.

"We had worked all year, worked hard," he said, "and it was an agreeable way of working; the brigade was very strong. We were already discussing how to renew our deal with each other. We were about to promise to work on for at least five more years when, suddenly, at the end of the first year, the farm administrators decided on their own to revise our agreement. They insisted we raise the payment for leasing. The economist hadn't changed. She was the same person who had been consulted on the original agreement. But she changed the clauses of the agreement," he said, staring and pointing at her. Then he looked directly at the chairman. I was moved by his poise. I didn't know why, but Loskutov was speaking candidly. Maybe he was speaking guilelessly. "That was really strange. We had paid 54 rubles per hectare, times 47 hectares. According to this new arrangement, we should suddenly pay 260 rubles per hectare—five times as much."

Horses haul freshly cut grass to cow barns

Farm Tudovik's bottling plant

I asked if he'd been given a chance to argue about what was fair, or to negotiate the increase. "They offered their new version—no discussion," said Loskutov. "We did not accept it. We waited a month and the brigade dissolved. We'd worked from May 1989 to May 1990, the entire first year of the agreement. Then we worked on into June. Nothing improved. We closed."

Had he felt a lot of envy and anger from the rest of the farm?

"Closing us down did quiet all the people," he said.

I asked if, before talk of closing came up, way back when they'd been working only six or eight months, others on the farm had wished to start similar lease brigades.

The zootechnician interrupted: "Interest in this kind of thing has been decreasing all over the country."

Chairman Drozhin spoke up: "A clause in the agreement said it could be revised each year and that it should be 'accepted by both sides.' But my analysis is that pushing this innovative lease-brigade form of enterprise on the collective farm made ordinary people fight it. In all enterprises I've encountered where lease brigades were introduced, we've experienced the fact that it's not proper. For some reason, participants think leasing and earning higher wages should be linked. But why? Leasing should not be a way of raising wages, but a way of raising the effectiveness of the enterprise! Some people don't want to work more effectively, just to earn more money." He seemed sincerely to see no connection between wage and ambition.

"Was that the case with Volodya?" I asked.

"No, it wasn't the case with Volodya," Drozhin said. "With him, the trouble was simply that other people on this farm were paid less, and I think that accounting was improper. No more questions about that..."

Nikolai awaited us outside. The whole conversation hadn't taken more than 20 minutes. Loskutov followed us to the jeep. He said he knew where Volodya lived down in Shuya and he would guide us there. We all tumbled in and excitedly took turns describing the conversation to Nikolai as we sped away.

"That economist!" Loskutov bellowed. "She whispered in the chairman's ear, told him to change the contract on us. Damn! We were doing so well, even with the worst cows—and we were still going up. Why did they stop us? He told you how they suggested we go to another barn with more cows? He didn't tell you it was to join a big regular work brigade, and we wouldn't have had any control over who did things. We'd just be workers there. He told you our barn was unhealthy? Our cows were healthy. They told us all that nonsense, and finally, we lost our enthusiasm. Not just us. There were other people on the farm who were thinking of doing the same thing—their enthusiasm got smashed, too."

Nikolai was soon laughing. "Is this how American journalists do their work? I like it. It does not happen here yet." Still, Nikolai remained a Party defender. "A leader like Gorbachev can introduce a new system. But if the farm managers won't accept it, it won't work. Managers have to support a leasing system and take steps that lead further on." Nikolai sounded pensive, but he was a deeply optimistic man. "You know," he said, smiling, perhaps at himself, "we could make our collective farm a corporation with shareholders. We may not be prepared to give over land to private ownership, but shares could be a step. A shareholder who became a pensioner could get income from prior work. A worker who moved away would have to sell back shares."

This was, in fact, what would happen to most collective farms two years later, after the Party fell. But the new system would leave the same personnel in charge, and they would still have the same problems with individual incentive and responsibility. Nikolai and I pledged that we would meet in another few years for the next chapter of the saga as we drove off to Shuya in search of Volodya. We had an hour to spare before my train left for Moscow.

Piloted by Loskutov, we drove into the back streets of Shuya, past factories and warehouses and anonymous blocks of flats. Opposite the police station, we turned into a pitted dirt parking lot. Loskutov disappeared inside a drab apartment building. The sun had set. Heavy clouds edged with pink hung in the sky. Rain splattered for a few seconds. Grandmothers gabbed and tiny children still played around the building. I stood watching. Volodya's neighbors stared back.

Loskutov dashed out. He'd found Volodya's wife, Olga. A neighbor would mind their two sons. She would come right down. She might be able to find Volodya. Olga ran toward us, breathless. She'd obviously just thrown on a sweater and dabbed spots of rouge on her cheeks. She was small and shy and seemed work-weary but eager. "Volodya spoke about your visit last year," Olga said. "He might be at his construction brigade's workshop—it's just a few blocks away." She crowded into the jeep, and we twisted through Shuya's back streets. We stopped before a ramshackle garage and Olga ran in. She came out shaking her head. "His truck headed up to Palekh a few minutes ago. He will be sorry."

"So are we," I said. "But let me ask—do you recall the moment he came home and said his rent was being raised and he'd have to quit farming?"

"He was angry. He yelled," Olga said. "They'd increased production. But then the new chairman raised the price on him, charged more for land, cows, everything. The agreement they were offered would have meant they'd be farming for no pay at all. Volodya said it was designed to make them quit. They weren't expected to sign it.

"If there was any way they could have continued," she said, "they wouldn't have quit. The herd of 200 cows they were supposed to get—that was phony, too. The men wouldn't have had freedom to work right. There were other bosses running that barn as a standard work brigade. And those cows were very bad, very old. Volodya was supposed to give back the herd they had improved and start all over again. It repelled them to contemplate going back to working with people they couldn't trust.

The established system rejects innovations the way a body eliminates foreign matter. Volodya's successful lease brigade challenged the system and proved it wrong, so he was wiped out.

"But the big reason may have been that so many farm residents were against them, and told them so openly. Donothings came by. 'We can do without you,' they said, and 'Your cows are hungry.' People came right into the barn and found fault with them.

"I believe you can't even blame those people—Volodya was doing something new. I blame the farm authorities: the new chairman and the farm economist. The chairman met the economist's demands. Nikolai Petrachkov could keep her in line, but the new chairman didn't. She was against Volodya from the start. They struggled to the last day, and right up until then he hoped to stay. She said, 'Stop farming.' The collective farm sent in men to close the brigade. They actually went in and moved the cows. Volodya learned something—he'll never start such a business again."

We drove fast and Yelena and I caught the night train to Moscow seconds before it left. After a while, the rocking train calmed us. Such rides had become intermissions. Yelena, whose graduate studies were in agricultural economics, stared out at the passing farms as the scenery faded into darkness. "It's too easy to say Volodya was kicked out by the economist and leave it at that," she said. "I think Volodya's problem wasn't the economist, or the new chairman. The problem is that up to now our people have rejected new structures. Volodya was too visible. Life on a farm is too open. In factories there are successful lease brigades. Our leaders thought the form would fit agriculture best, and it hasn't. Those six or eight workers set a pattern that might have spread and taken away the jobs of half the villagers."

At 5:30 in the morning, the train groaned into Moscow. I caught a few hours sleep and that evening had supper with Viktor Lishchenko, the reform agricultural economist who'd pointed me toward Ivanovo. He was more outspoken than usual. He said he was not surprised to hear that Volodya had been thwarted. "The system we evolved rejects innovations the way a body rejects foreign matter," he said, "which has caused our descent in the world economy. Our country is full of innovators, but the system rejects them all. Volodya's successful lease brigade was a showcase that proved the system is wrong. So he was wiped out. When economic adventurousness is permitted in the USSR, that muddled reach for progress still retains centralized management, leading again and again to the same fate for energetic, creative people." His words made me wish that I'd met up with Volodya again. I still wanted to know how this aspiring entrepreneur had fared.

Two years later, in 1992, I was back in Russia, but much had changed. The previous August an abortive coup had led to the fall of the Communist Party and to Gorbachev's demise by year's end. I'd waited a season for the smoke to clear, then returned to someplace new. I was no longer the "American Delegation" when I traveled, no longer an official guest of official institutions, bound by official itineraries. I found my own translator, a dissident writer and historian named Sergei Sossinsky, who'd spent his teen years in Queens, New York, because his father had worked at the United Nations. We stayed not in hotels as party guests but on couches in the apartments of friends of friends, roughing it.

Volodya had become a poised, worldly trader, Russian style—a wheeler dealer. I felt downright flummoxed that I'd taken this adept businessman for an earnest hayseed.

Sergei and I drove back to Shuya. We stood in Volodya's apartment house courtyard, where Loskutov had led us in 1990 and where I'd chatted with Volodya's wife. And there, as if materialized from a dream, stood Volodya, waving. Finding just the person I wanted to see, after so many controlled trips, had proved straightforward. Without political escorts, we'd just gone and done what we'd set out to do.

I remembered Volodya's puzzled, lopsided smile the moment I saw it again. He'd filled out. He'd prospered. His apartment was furnished with leather couches and beautiful Oriental rugs—not what I'd have expected of the hardworking would-have-been farmer. Fine icons hung on the walls. His wife served us liver, which was costly and "fresh from the market," she said—a lavish gesture of hospitality. She carried out tea and Metaxa brandy, sausages and bread, Swiss chocolates and cakes. "My life has changed so much," he said. "I'm another Volodya now."

And he was another Volodya. He'd struck me during our earlier visit as more directed than the average farm worker, but I'd missed his eloquence. He recalled the events of May and June 1990 that led to the end of his lease brigade: "The culprit who ruined my farm was Drozhin, the replacement chairman—not the chief zootechnician nor chief economist, although they also didn't like what we were doing. Drozhin was the kind of chairman who got irritated if he merely heard that something new was going on. We had worked hard the whole winter. The cows milked well. They all got with calf. And then the new calves were born, and none died; we'd given the farm excellent new cows. We'd replaced a quarter of our herd, and we were ready to go on working.

"There was a big meeting. The collective farm board came. Drozhin came. The city Party committee came. Sorokin, the big boss, even came, although he didn't speak. All Farm Trudovik's board members opposed us. But they accused us of just one thing—getting paid well. I told them, 'In that case, we will just throw everything away and start drinking like your regular collective-farm workers do. Then you'll be happy, because our wages will go down.' I was angry.

"Near the end of winter, the collective farm just began dumping cow manure into our fields one day. They didn't spread it, just dropped truckloads of it. I asked, 'Why do you want to ruin our fields?' The field supervisor said, 'You won't be working on them anyhow.' I went over to Nikolai Petrachkov at Farm Gorky. He's a very serious man, the one I still respect among the area's chairmen. Couldn't I move with my crew down the road and join him again? He was friendly. But he said that he didn't have room.

"Soon Drozhin said if we didn't get out of the barn, they'd simply flatten it with a bulldozer and kill the cattle. I'd taken care of those cows for a year. I felt sorry for them. I yelled at him, 'You're a Communist—you should be thinking about the workers' prosperity, but instead you're thinking of murdering a good herd of cows. How can a farm chairman say words like that? What kind of master of a farm are you?'

"I drove right to Ivanovo and saw the deputy Party chief of agriculture in a big office. I told my story and got all excited. He pretended to listen, made a show of writing something down, and that was the end. I worked exactly a year, May first to May first, International Workers' Solidarity Day—but solidarity with whom? They closed up our barn and drove the cattle back to a bigger barn in Putich. Someone told me that along the way 12 cows stopped and drank water into which ammonia fertilizer had fallen. They died. Even after that, I saw Drozhin once more and asked if he'd sell me the farm—not the land, just the farm building and a small plot. He said I didn't have enough money. I told him I had enough to buy his whole damned collective farm. I was pissed off, but after that I just went away."

If I'd missed Volodya's intensity and eloquence, I'd also underestimated his ambition and commercial skill. He'd worked hard at the lease brigade because he'd seen it as an opportunity to make good money. It had failed, but just as Volodya had spotted the chance to farm under Nikolai Petrachkov, he went about finding the next opportunity.

"I have a family, kids. I needed to work. It was still *perestroika*. Private cooperatives were suddenly allowed at that point. So we started a small construction cooperative. We repair schools, nurseries, invalid homes—only public buildings. We pay our workers good wages. So far, the state's reimbursements to us for our work are coming through OK.

"The enterprise also now has a 'commercial section.' We're middlemen—I'm not ashamed; a middleman takes risks. Our laborers did maybe 300,000 rubles' worth of repairs at a cloth mill. In exchange we got textiles worth that. We trucked them to Voronezh, where there is not much fabric, but there's a candy factory. For each kilo of candy that sells in stores (if you can find any) for 120 rubles, we traded 60 rubles' worth of fabric. We sell the candies here for 80, 90 rubles.

"Our business is growing, but not as fast as it could. The country's falling apart. Someone else wants us to trade textiles for grain, which would be attractive—there's not enough bread in Ivanovo. But we'd have to make a preliminary payment, and we can't because the banking system has blown up. People line up at banks, and then the bank gives them only 500 rubles apiece from their own accounts. We have other troubles. Some companies that approach us are phonies, and we can't tell. There's absolutely no basis for mutual trust. And consumers aren't buying things, because they have no cash around." This was Volodya the plain dirt farmer giving me the lowdown. I was amazed.

"For 70 years our government tried to invent new economic laws. But you can't. The world lived according to definite economic laws already. We are making the transition back to what the rest of the world knew—provided the government isn't taken over by people looking backward. There is a danger—maybe not from Moscow, but from the provinces, where there are millions of workers who remember Brezhnev's era as a good time. It's so hard now for them to live."

"I'd like to produce something, not just be a middleman. But it absolutely can't be done now. I wanted to buy a brick factory, but land isn't being sold. You still can't get to own the land under a factory, never mind what Yeltsin decrees. I'd like to start a quarry. I know that the state's cost for producing bricks is very high. State structures would be no competition for us. I figured it out.

"I am also trying to start making flooring materials. I met a Pole in Moscow who was in a joint venture with Germans producing parquet flooring—high-quality products. There are empty warehouses just outside of town. We have equipment. We went to every official. But they wouldn't let us rent or buy a place to put the machines. You can't set them down in the open air. You need quarters, and land. We finally got space as part of a state enterprise, but I'm sure they'll soon see we're prosperous and keep cranking up the rent. They'll add their workers in with ours. They'll demand a share. We'll see. I'm just doing what I'm doing. I'm doing pretty well." He laughed. It was a businessman's excited laugh.

And then Volodya showed yet another side of himself: "All my life I've been fascinated by icon paintings. Tsar Peter the First visited here once to worship a certain icon. I almost went to jail for my interest in icons, so I stopped. But now you can do anything. Look at these icons—sixteenth century, nineteenth century. The Communists destroyed an enormous number. Now they're getting harder to find. I travel around. People bring them to me, and I buy them. I paid over a thousand dollars for this," Volodya said, handing me a small icon of a doleful Mary. "I knew what I was doing. I confess—I'd go to Moscow and sell them. Russian customers paid a lot—they knew the good ones. Italians fell in love with them and showed it. But the Americans wouldn't show emotion. They'd just wander back three hours later and act cool.

"I have no scruples about the way I make money myself—except for criminal things. I've worked so hard that I have hardly been home for the years since I lost the farm. I'm slowing down a little now, but never on weekends. Then, I'm always in Moscow doing business."

I felt downright flummoxed that I'd taken this adept businessman for an earnest hayseed. He had become a poised, worldly trader, Russian style. A wheeler-dealer. That a black-market icon trader had seemed to be a man disposed toward constructive action and clear sight was a measure of the confusing moral context that Soviet power had spawned. They'd banned property and religion, and they'd stifled art, and Volodya had found a bold way to traffic in all three. I took it as a sign of a lively spirit in a tough spot.

My farmer of the future had landed on his feet. In fact, he'd been nimble all the while, eyes fixed on the opportunity at hand. How innocently I'd read him, in 1989, sitting in the hay shed, talking quietly so my keeper, Lev, wouldn't overhear Volodya's criticisms of Soviet farming. Volodya made more sense this way. A man who'd sought out the opportunity to farm independently, years before most citizens even understood the possibility, would notice other rewarding outlets for his industriousness. "The effects of bad laws in Russia are lessened by their bad enforcement," Marquis de Custine wrote a century and a half ago.

Over the next few days, I visited farmers who, like Volodya, had stumbled off on their own agricultural initiatives. A few were still in business. I was hunting for a success—a replacement farmer of the future, now that Volodya had moved on to commerce and the building trades. But I never found one. These young men limped along in the midst of local hostility, still as hemmed in by bureaucrats as ever—bureaucrats who just wanted to hang on to their own dominions. These individual farmers were still denied credit, supplies, machines, and markets. They were going nowhere, and would not go anywhere soon, although their nation would continue to need their sort of help. ■

The author in Red Square

WHARTON REPROGRAPHICS

THE FOREIGN EXCHANGE MARKET: BACKGROUND NOTE AND PROBLEM SET

International finance plays a large though often hidden role in all of our lives. Whether we are driving to work in a Japanese car, wearing clothes tailored in Hong Kong, or having a German beer before dinner, we are using products originally purchased with currencies traded through the foreign exchange market. An understanding of this market is essential to firms that operate globally because they are exposed to foreign-currency fluctuations. Moreover, the market has grown substantially in recent years because of floating exchange rates, deregulation, and the advancement of communications technology. Though it is difficult to measure the size of the foreign exchange market, in 1985 the Group of Thirty research organization estimated the *daily* foreign exchange turnover at $150 billion. By 1986, however, a survey of the central banks from the United States, the United Kingdom, and Japan reported that daily turnover in the top three centers alone was $188 billion. Continued reform will make the foreign exchange market an increasingly innovative and exciting arena in which to participate.

MARKET PARTICIPANTS

The foreign exchange market involves participants buying and selling currencies all over the world. Elaborate communications systems, including telephones, telexes, computers, and news wires, link individuals worldwide. Trading takes place throughout most of each day, starting in Tokyo; moving west to Singapore and Hong

This case was prepared by Research Associate Richard P. Melnick and Professor W. Carl Kester. Copyright © 1987 by the President and Fellows of Harvard College. Harvard Business School case 287-033.

Kong; continuing to Zurich, Frankfurt, and London; and then ending in New York, Chicago, and San Francisco.

Participants in the market include individuals, corporations, commercial banks, and central banks. Each has different motives for transacting. Commercial banks with large foreign exchange sales and trading departments act as dealers when they buy and sell foreign exchange for clients. Their clients, usually corporations, need to transfer purchasing power to and from other countries in order to buy foreign goods and services and to invest in foreign assets. Commercial banks profit from the spread between buying currencies low at a "bid" price and selling them at a higher "ask" or "offer" price. Individual speculators profit when there is a change in general price levels and they have predicted the change better than the market. Individuals, including tourists, importers, and exporters, use the foreign exchange market to conduct trade or investment transactions.

Central banks have two official reasons for participating in the foreign exchange market: (1) to monitor the market and (2) to intervene for purposes of policy implementation. A central bank might intervene, for example, to smooth out extreme currency fluctuations or to meet obligations under formal agreements such as the European Monetary System (EMS).[1] A striking example of the central banks' ability to influence the market was the Group of Five's effort to depreciate the dollar in September 1985. Through the sale of billions of dollars, the dollar depreciated 35 percent in 10 months. Intervention on this scale was unprecedented.

Profiting on their currency reserves has increasingly become an additional motive for some central banks to enter the foreign exchange market. The Soviet Union, acting through its Vneshtungbank trade bank, reportedly has been so profitable in its foreign exchange dealings that many suspect information leaks from the U.S. Commerce Department and the German Bundesbank.[2]

Central banks can act discreetly or conspicuously, and they often exploit the market power derived from their mere threat of action. The Bank of England is reported to be the most professional bank and has cleverly used the market to discourage speculation against sterling. The German Bundesbank is considered sophisticated, but with a somewhat tarnished reputation as it has been thought to give domestic banks advantageous information. By telling their national banks outright what to do, the Bank of Japan acts aggressively without trading directly. Traders generally like to avoid the U.S. Federal Reserve Bank, whose intervention is often thought to be clumsy and inept.[3]

---

[1] The European Monetary System is an organization composed of all the European Economic Community members. It was founded in 1979 to stabilize exchange rates among the European countries. Their currency basket, called the European Currency Unit (ECU), is composed of a predetermined quantity of each currency, where the components reflect each member's relative size. Each nation's currency may not deviate from the ECU by more than 2.25 percent (the British pound does not follow the ECU restrictions), but the ECU itself floats against the rest of the world.

[2] Henry Sender, "Games Central Banks Play with Currencies," *Institutional Investor*, November 1985, pp. 100–110.

[3] Ibid., p. 108.

A 1986 survey of central banks in the United States, the United Kingdom, and Japan shows that London, New York, and Tokyo, with $90 billion, $50 billion, and $48 billion, respectively, in estimated *daily* turnover are the leading foreign exchange centers. London is the world's largest foreign exchange center because it has several advantages: history, geography, and a healthy regulatory environment. Expertise and tradition go back to the days when the pound was the most widely traded currency. More important, however, is the fact that London can trade with Tokyo and Hong Kong in the morning and New York in the afternoon, as well as Frankfurt and Zurich continually. Trading has increased substantially since 1979, when foreign exchange regulation decreased in Britain and banks were permitted to trade directly rather than through brokers.

Tokyo's foreign exchange market has grown dramatically from an estimated daily turnover of only $8 billion in 1985. Under pressure from the United States, the Japanese have also decreased market regulation and have thereby surpassed Hong Kong and Singapore in average daily turnover.

New York's foreign exchange system grew during much of the early 1980s, but suffered under the weak dollar of 1986 and the increased power of European central banks. Nonetheless, the greatest volume occurs in the European afternoon when New York is open. Because of the deep market, this is the safest time to execute large transactions.

THE SPOT MARKET

The *spot* exchange rate allows for the buying and selling of foreign exchange with settlement in 2 business days, known as the *value date*. According to the aforementioned central bank survey, spot transactions accounted for 73 percent of all business in London and 63 percent in New York. Spot rates are quoted in terms of both currencies in *The Wall Street Journal*, one just being the reciprocal of the other. On September 2, 1986, for example, the spot rate for Swiss francs (SF) against U.S. dollars was SF = $0.60994, and the cost of the U.S. dollar was therefore 1/0.60994 or 1.6395 Swiss francs.

Quotes are said to be direct or indirect. A *direct* quote is the price of a foreign currency unit in terms of the home currency. According to the previous example, the direct quote for Swiss francs in the United States was $0.60994. An *indirect* quote is the price of a unit of home currency measured in the foreign currency. Thus, the indirect quote for Swiss francs in the United States was SF1.6395.

With this system, a quote's definition depends on the country of reference. To clarify this ambiguity, New York bankers established the *European terms* convention, which uses the U.S. dollar as the common denominator. Thus, traders in the United States, Switzerland, or any other country would all quote Swiss francs as SF1.6395/$. The exceptions to this rule are British sterling and the currencies of several former British Commonwealth countries including Australia and New Zealand. The exceptions exist because the British pound (£) was formerly not a decimal currency and it was easier to quote sterling in terms of other currencies. Today, these currencies are still quoted around the world in "American terms," which is the number of dollars per unit of their currencies.

Cross Rates and Triangular Arbitrage

If a Swiss bank wants German marks, it could get a rate of Swiss francs per deutsche mark from a German bank. From the perspective of dealers in a dollar-based market, this quote would be a *cross rate*. These rates can be given with either currency serving as the denominator and they are most common among the Western European currencies. These cross rates are typically used by central banks of countries in the EMS trying to determine when they need to intervene.

Cross rates should equal the rates resulting from conversion to and from U.S. dollars (except for small differences attributable to transaction costs). Before using a cross rate, it is useful to see if transacting through dollars as an intermediate currency yields a different rate. For instance, if a dealer offers a dollar for 6.6625 French francs (FF) or 2.0320 deutsche marks (DM), then the cross offer rate should be 3.2788 French francs per deutsche mark. With bids of FF6.6575 and DM2.0310 per dollar, the corresponding cross bid rate would be 3.2779. If the cross bid rate were actually 3.3000 French francs per deutsche mark, traders could engage in *triangular arbitrage* to devalue the cross rate and push the system into equilibrium. Traders would take the following three steps: (1) Sell DM1,000,000 (the "overpriced" currency) for French francs at a cross rate of 3.300 French francs per deutsche mark and receive FF3,300,000; (2) sell French franc holdings against dollars at a spot offer rate of FF6.6625 per dollar and receive $495,310; and (3) sell U.S. dollar holdings against deutsche marks at a spot rate of 2.0310 deutsche marks per dollar and receive DM1,005,975. The potential for this profit of DM5975 would erode quickly as deutsche marks are sold against French francs at the overvalued cross rate.

Newspapers usually give only the bid rates or estimates at the midpoint of dealer spreads under European terms (except for sterling), while actual dealers give buying and selling quotations. "Offer" or "ask" rates under European terms will be higher than bid rates since dealers will want to sell dollars at a higher price than that at which they buy dollars. The difference between the two rates is usually quite small. For example, Citicorp could buy French francs (sell dollars) at FF6.6675 per U.S. dollar and sell francs (buy dollars) at FF6.6625 per U.S. dollar. The spread is FF0.0050 or 50 points, where a "point" refers to the quote's last digit.[4] Though the bank's spread might be as small as 0.0005 of the transaction value, this is how banks profit in the foreign exchange market. With a more typical 10-point spread, bank revenues on $1 billion of transactions could be $1 million. The spread travelers experience can be 10 times as high because executing their many small transactions requires higher inventory carrying costs and virtually the same amount of paperwork as that of large transactions.

---

[4] A "point" or "pip" can refer to two, four, six or any number of decimal places. If the Japanese yen were quoted in American terms with an ask of $0.001235 and a bid of $0.001230, this would still be considered a 5-"point" spread.

THE FORWARD MARKET

When business deals require individuals and corporations to pay or receive foreign currency in the future, they often prefer to transfer the risk that currency values will change. The *forward market* allows corporations to establish the exchange rate between the two currencies for settlement at a fixed future date. This is known as *hedging* or *covering* foreign currency exposure. The *forward rate* is fixed at the time the contract is made, but payment and delivery are not made until the value date. Forward contracts are often negotiated against dollars, though European traders also exchange domestic European currencies against each other. The most frequent forward contracts are for 1, 2, 3, 6, or 12 months. Banks actively trade in these markets both for their customers and for themselves. "Odd-dated" forward contracts (i.e., maturities other than the most common ones) are possible, but they are more expensive because of the thinner market. Long-dated forwards and special quotes can be arranged for up to 10 years with major currencies.

The ratio at which two currencies will be exchanged on a future date is the *outright* forward rate. These rates are usually printed in newspapers and are used by banks when dealing with their clients. Foreign exchange traders like to abbreviate, however, and quote forward premiums or discounts in terms of points.[5] A *dollar premium* implies that forward dollars are at a premium over spot, and the French, for example, need more French francs to buy dollars forward than to buy dollars spot. Similarly, a *dollar discount* means one needs fewer units of foreign currency to buy dollars forward than to buy dollars spot.

In order to tell if a currency is at a premium or a discount, compare the size of the forward bid points with the forward ask points for a given maturity. If the forward bid points are higher than the forward ask points, the U.S. dollar is at a forward discount. For example, if the 3-month forward points for deutsche marks are 38/35, then the forward bid points of DM0.0038 are greater than the forward ask points of DM0.0035 and the forward market is at a dollar discount. Similarly, if the 3-month forward points on French francs are 50/55, the bid is smaller than the ask, and the dollar is at a forward premium.

To calculate the outright forward rate, one needs to know if the U.S. dollar is at a premium or a discount. Looking at the forward bid-offer points, if forward bid points are greater than the offer points, the currency is at a U.S. dollar discount and the trader should *subtract* the forward bid-offer points from the quoted spot rates. If the forward bid points are less than the forward offer points, the currency is at a dollar premium and the trader should *add* the forward points to the respective spot points.

---

[5] Dealers often do not use the decimal point for forward bid-offer points. Whether talking on the phone or placing quotes on a screen, traders give the complete first bid of the spot quote (e.g., FF6.6575). All that is given for the ask, however, are the digits that differ from the bid (i.e., 75). A slash (/), communicated orally by the word "to," separates the bid and ask quotes. Thus a spot quotation for the French franc might appear on a screen as 6.6575/625, implying an offer rate of 6.6625, and the corresponding 3-month forward rate would be represented as 263/296. One adds these points to the *last* three digits of the spot rate to get the outright forward quote.

Forward quotes can also be presented as a percent-per-annum (% p.a.) deviation from the spot rate. By convention, the formula for calculating forward premiums and discounts as percentages using European terms, with n equal to the number of months in the contract, is:

$$\text{Forward premium or discount (\% p.a.)} = \frac{\text{spot rate} - \text{forward rate}}{\text{forward rate}} \times \frac{12}{n} \times 100$$

Using spot and 3-month forward quotes for deutsche marks of 2.2605 and 2.2406, respectively, we would have:

$$\frac{2.2605 - 2.2406}{2.2406} \times \frac{12}{3} \times 100 = 3.55\%$$

When using American terms, the formula changes as follows:

$$\text{Forward premium or discount (\% p.a.)} = \frac{\text{forward rate} - \text{spot}}{\text{spot}} \times \frac{12}{n} \times 100$$

Thus, in the above example we would have

$$\frac{0.4463 - 0.4424}{0.4424} \times \frac{12}{3} \times 100 = 3.53\%$$

This differs from the percentage using European terms only by a rounding error.

Covered Interest Arbitrage

Many factors explain the changing values of foreign currencies relative to the U.S. dollar. Spot rates can be quite volatile, influenced by news, rumors, speculation, supply and demand imbalances, and central bank intervention. Over the long run, spot rates are affected by relative interest rates and inflation rates, trade imbalances, and purchases or sales of foreign assets.

Perhaps the dominant factor influencing the forward rate of a currency relative to its spot rate is the difference between nominal interest rates on foreign currency–denominated investments and comparable U.S. dollar investments. Investors will compare interest rates around the world in pursuit of the best possible return for a given level of risk. Large banks and corporations seeking short-term use of cash will generally look at Eurodeposit rates since such deposits are quite accessible, secure, and competitively priced because of the less-regulated nature of the Eurodeposit market.

However, the highest nominal Eurodeposit rate does not necessarily imply the highest expected yield in home currency terms. If the currency in which the deposit is made depreciates against the home currency, then the actual yield on the deposit in home currency terms will be less than the quoted nominal yield. For example, a 1-year Eurofranc deposit may offer a yield of 11⅝ percent compared to a 1-year Eurodollar deposit offering 7⅞ percent. But if the franc depreciates by 3.36 percent or more, the interest rate difference favoring the franc will have been completely offset from a U.S. dollar perspective. The depositor could hedge this risk by selling francs (purchasing dollars) forward at the prevailing market exchange rate, thus locking in the future rate at which the principal plus interest can be converted to dollars. In principle, assuming an efficiently operating foreign exchange market, the prevailing 1-year forward contract rate for French francs will reflect the market's collective judgment about what the spot rate will be 1 year later. In other words, the current forward rate should be an unbiased predictor of the future spot rate, a condition known as *forward parity*.

Because money can flow quickly and in large volume from one Eurocurrency deposit to another, and because forward rates may be thought of as unbiased predictors of future spot rates, Eurodepositors aggressively seeking the highest *effective* return (including expected exchange-rate movements) should drive Eurodeposit rates, spot exchange rates, and forward exchange rates into an interdependent relationship in which expected yields are identical across currencies. This condition is known as *interest rate parity*. It can be expressed notationally by the following equation:

$$F = S\frac{(1 + R_F)}{(1 + R_\$)}$$

where F = the forward exchange rate expressed as units of foreign currency per dollar (European terms)

S = the spot exchange rate expressed as units of foreign currency per dollar

$R_\$$ and R_F = the Eurodollar deposit rate and the other Eurocurrency deposit rate, respectively

If this condition is violated, a *covered interest arbitrage* opportunity will present itself (for this reason, the above equation is also frequently referred to as the "covered interest arbitrage" condition). For example, consider the terms presented in **Table 1**.

The French franc is selling at a forward discount against the dollar. Using the midpoints of the quotes, the discount is 1.88 percent on an annualized basis. This is

TABLE 1

Exchange Rate and Interest Rate Quotes	Spot	3-Month forward
French franc	7.7150/200	300/425
3-month Eurofranc rates	11⅝–11¾%	
3-month Eurodollar rates	7⅞–8%	

smaller in absolute terms than is theoretically justified on the basis of differences in interest rates:

$$3.68\% = \{[(1 + 11.6875/400)/(1 + 7.9375/400)] - 1\} \times 4 \times 100$$

Consequently, an arbitrageur could profit risklessly by taking the following steps:

1. Borrow 1,000,000 Eurodollars for 3 months at 8 percent.
2. Purchase FF7,715,000 at the spot bid rate.
3. Invest the francs for 3 months at the Eurofranc deposit rate of 11⅝ percent.
4. Simultaneously sell francs (buy dollars) forward at the forward offer rate of 7.7625. The amount of the forward sale should equal the total amount of franc proceeds at the deposit's maturity, or FF7,939,217 [FF7,715,000 × (1 + 0.11625/4)].

In 3 months, the arbitrageur would receive the Eurofranc deposit with interest, execute the forward contract, and obtain $1,022,765 (FF7,939,217/7.7625). This could be used to repay the Eurodollar loan with interest amounting to $1,020,000 [$1,000,000 × (1 + .08/4)], leaving a gain of $2765. Assuming total transaction, brokerage, and cable cost of 0.25 percent, or $250 on the $1,000,000 deal, a net profit remains of $2515. Note that this was obtained risklessly and without any equity directly committed by the arbitrageur. As many people try to exploit this opportunity in volume, interest rates and exchange rates will adjust until parity is restored and riskless profit opportunities are eliminated.

INTERBANK TRADING

The vast bulk of foreign exchange transactions takes place in the *interbank market*, where banks and foreign exchange brokers determine exchange rates. Dealers try to develop a sense for where the market is going and then buy or sell currency either for their clients or for their own portfolios. After dealer A decides what he wants to buy or sell, he calls foreign exchange dealers at other banks and "asks for the market." He wants to know whether the dealers' rates are competitive and does not say whether he intends to buy or sell. Foreign exchange dealer B (being called now) must use her intuition to determine what caller A wants to do before offering a quote. Many techniques are used in an effort to get the best rate. For example, if the first caller wants to buy deutsche marks, he might first ask for the market in pounds to confuse the dealer and only later ask for deutsche marks.

Aside from figuring out what the caller wants to do, dealer B must also devise a strategy that meets her goals. If dealer B wants to *sell* deutsche marks (buy dollars), she might present a bid of DM2.5235 per dollar, 2 points higher than other banks', which are at 2.5233. The dealer must also quote an offer, and this should not be better than the market since she does not want to own deutsche marks. Convention holds that the caller can choose either rate and any amount within 1 minute after the quotes. After the minute, dealer B can change her quotes. When a difference in the fourth decimal place can represent thousands of dollars on a big transaction, good judgment and fast thinking are vital.

The parties exchange confirmation papers after a deal has been arranged over the phone. Settlement occurs on the value date as the currencies are exchanged through the banks' clearing accounts. Client rates are generally determined in the retail bank draft market, while the interbank rate is determined in the wholesale market. The largest corporate clients get rates similar to those offered to other banks.

Sometimes banks cannot make markets among themselves. When a private deal cannot be arranged, foreign exchange brokers are used as intermediaries. Banks call brokers and tell them how much foreign currency they want to buy or sell and at what rate they will do the transaction. The broker then takes many requests to buy or sell and tries to find banks with complementary goals. If two banks' exchange-rate terms can be matched, a trade is made. The banks only learn of each other's identities after a contract is settled. Following the contract's consummation, the broker receives a fee from each bank. Banks always prefer to make their own market since the broker's fee makes the same transaction more expensive.

OTHER FOREIGN EXCHANGE INSTRUMENTS

As indicated, the foreign exchange market allows individuals and corporations to transact in the spot market for immediate delivery and in the forward market for settlements in the future. Other foreign exchange vehicles and markets are also appearing. Although they accounted for less than 1 percent of total foreign exchange transaction volume in 1986, their importance is likely to grow.

While forward rates are quoted by commercial banks, *foreign-currency futures* are homogeneous contracts traded in physical marketplaces such as the International Monetary Market (IMM) in Chicago. Buyers and sellers of foreign-currency futures deal through the exchanges rather than directly. Advantages of these contracts are that they are relatively small and, because of their homogeneity, are highly liquid. However, homogeneity also implies inflexibility, and there are high costs on large futures transactions.

Though complicated, expensive, and relatively illiquid, *foreign-currency options* are becoming an attractive transaction alternative. Options allow a company to hedge against foreign exchange losses, but they also provide opportunities for bold speculation in the market. Unlike futures contracts, options represent the right, but not the obligation, to buy or sell a specific amount of currency at a fixed price on or before some future date. Since the option holder can always avoid exercising the option if it is uneconomical to do so, options make possible large upside gains while limiting downside losses. Costing an average of 5 percent to 6 percent of the underlying contract amount, options are expensive compared to the spreads on interbank forward contracts. Though options are not for everyone, more companies are using them to make a profit and hedge risk in times of volatile exchange-rate fluctuations.

A common tool used to hedge foreign-currency risk is a *foreign currency-swap*. A foreign-currency swap allows one party to swap principal and interest in one currency for principal and interest in another. Usually the participants borrow different currencies, exchange the receipts, and then service each other's debt. This arrangement eliminates exchange risk for the transaction and makes future currency fluctuations irrelevant. Sometimes a domestic company is able to borrow funds at a cheaper rate than a foreign company could. Swaps provide cost savings when two companies with complementary needs and desires get together. The World Bank is a frequent participant in these swaps and arranged the first important swap with IBM in 1981. Interest rate swaps, developed in 1982, allow a company with a fixed-rate asset or liability to convert the security into a floating-rate obligation or vice versa. Since then, foreign-currency and interest rate swaps have typically been used together. Further details concerning swaps are provided in the *Note on Foreign Currency Swaps*.

EXCHANGE-RATE PROBLEMS

The mechanics of exchange-rate arithmetic, though fundamentally simple, can be confusing for those not using it on a regular basis. The questions below are designed to provide practice in the more common manipulations. They are of low to moderate difficulty with the easiest problems occurring first. **Exhibits 1** through 3, along with the general background information provided in this case, are sufficient to solve the problems.

1. Are the dealer quotes shown in **Exhibit 1** direct or indirect? If DM1,000,000 were sold spot, how many dollars would be received? When would settlement normally take place?

2. Examine the cross spot rates shown in **Exhibit 2**. Are there any triangular arbitrage opportunities among these currencies (assume deviations from theoretical cross rates of 5 points or less are attributable to transaction costs)? How much profit could be made on a $5 million transaction?

3. What would be the $/SDR (SDR = U.S. dollars per unit of currency) bid if the SDR appreciates 15 percent against the dollar? What would be the *SDR/$* offer rate if the SDR appreciates 15 percent?

4. Which currencies are at a dollar discount and which are at a dollar premium? What are the outright forward rates for the pound? For the French franc? Using the midpoints of bid-ask spreads, what are the forward premia or discounts on an annualized percentage basis for both these currencies?

5. A private speculator expects the yen to depreciate 7 percent against the dollar over the next 3 months. How can the speculator try to profit on these expectations through (a) spot market transactions only, and (b) forward market transactions only (assume no margin requirements or restrictions on transactions in credit markets)? What will be the expected dollar profit on a $1 million position in each case? What other considerations should factor into the speculator's choice between the spot and forward markets for purposes of profiting on future movements of the yen?

6. A U.S. corporate treasurer will receive a £2 million payment in 30 days from a British customer. The treasurer has no strong opinion about the direction or magnitude of changes in the sterling spot rate, but would like to eliminate the uncertainty surrounding such movements. Within the context of the rates shown in **Exhibits** 1 and 3, what options are available to the treasurer for hedging the foreign exchange risk associated with the sterling payment? What is the expected cost (expressed as an annualized percentage) of each alternative? Which alternative should the treasurer pursue? How would your answers to the above questions change if the treasurer believed very strongly that sterling would trade at $1.45?

7. Compare the 1-year forward premium or discount on the French franc to the 1-year Eurodollar and Eurofranc interest rates shown in **Exhibit** 2. How can this situation be arbitraged?

8. Examine the spot rate and the 6-month forward rate for the British pound. Suppose a speculator anticipates that the pound's spot rate and the 6-month Eurodollar deposit rates will be unchanged from their present levels in 6 months' time. However, at that future date, the 6-month Eurosterling deposit rates will have changed to 10.0000 percent to 10.0625 percent per annum. What should be the new 6-month forward rate for the pound if covered interest arbitrage opportunities are to be avoided half a year from now? How can the speculator profit from the expected change in the interest rate difference while remaining in a "square" position (i.e., offsetting foreign exchange purchase contracts with sales contracts) at all times? What will be the expected dollar profit per pound? How will this expected profit change if the spot rate 6 months later does not remain constant but changes to 1.5500/10? To 1.4500/10? What circumstances might cause the speculator to realize a loss rather than a gain?

55

LIPPINCOTT LIBRARY
DOCUMENT DELIVERY

WHARTON REPROGRAPHICS

Retail Foreign Exchange Trading in Prague: Are Tourists Rational?

ROBERT M. DUNN, JR.*
George Washington University
and
CERGE, Charles University

For more than two centuries economic theory has been based on the core assumptions that economic agents maximize their self interest and that they are rational. These assumptions of avarice and ingenuity have sometimes made economics unpopular among scholars from the humanities, and have begun to be questioned even by economists. A significant literature has developed on altruistic economic behavior, and a new sub-field of economics, socio-economics, has raised additional questions about whether people always maximize their self interest in a rational manner.[1]

Economists have defended the assumption of rationality with more vigor than that of self-interest, and have used appeals to differing tastes to explain apparent contradictions to that assumption. If what appear to be the same products are sold at two different prices under what seem to be the same conditions, minor or even tiny differences are noted so that those paying the higher price can be viewed as being rational but as merely have tastes for the more expensive product. The costs of gathering price information are also used as an excuse for people paying a higher price for the same good. If the implicit wage on the time required to investigate alternative prices exceeds the likely savings from finding the lowest price, consumers can rationally pay more for the exact same product than the price at which it is available from another vendor.

---

*Direct all correspondence to: Professor Robert M. Dunn, Jr., Department of Economics, The George Washington University, 2201 G. Street, N.W., Washington, DC 20052.

Journal of Socio-Economics, Volume 26, No. 5, pp. 469–473
Copyright © 1997 by JAI Press, Inc.
All rights of reproduction in any form reserved.
ISSN: 1053–5357

Retail foreign exchange trading by tourists makes it possible to revisit this issue by observing the exact same product being sold at the same place and time at easily discovered prices to what can be assumed to be rational adults.

During the last 7 years, Prague has become an enormously popular tourist destination, and the city consequently has hundreds of retail foreign exchange trading offices, which are sometimes labelled as cambios - weschels. Americans and other tourists typically arrive in Prague with travelers checks denominated in their home currencies and need to exchange them for Czech crowns. The exchange offices provide the exact same product (crowns in the form of paper currency) in exchange for commissions which appear to vary widely. These offices are often in close proximity; more than a dozen can be found on or immediately adjacent to Wenceslas Square, a wide avenue of about three blocks in length.

The exchange rates at which transactions are to occur are typically posted conspicuously, but the commission percentages, which are applied to those rates by various offices, are sometimes shown in small letters at the bottom of the sign illustrating the exchange rates. Information on the commissions, however, is posted so it can hardly be claimed that consumers are unable to discover the relevant prices. Rational consumers (at least those undertaking more than one transaction while in Prague) could be expected to understand the commission system and to avoid those trading offices where the combination of the exchange rate and the commission produced significantly fewer Czech crowns for a $100 travelers check than were readily available within 100 yards on the same street. Simply put, exchange rates adjusted for commissions ought to be arbitraged together, or at least very close together, within a small areas such as Wenceslas Square or the other popular tourist areas.

Tourists might pay more than was necessary once, but they could certainly be expected to avoid this outcome a second time and to communicate their experiences to others in their family or group, so the similar losses could be avoided in the future. Exchange offices with particularly high commissions should be driven out of business by those offering more attractive outcomes within the same immediate area.

The assumption of the rational consumer (tourist) does not do well in the retail foreign exchange market in Prague. The exchange rates vary only slightly among the offices, but the commissions are very different. In the immediate Wenceslas Square area, commission rates as high as 9.8% and as low as 1.5% were found in a brief period on August 28, 1996. Combinations of exchange rates and commissions were discovered during a half hour on that day which produced as few as 2306 crowns and as many as 2511 crowns for a $100 travelers check, which is a range of 8.9%. During that day the crown/dollar interbank rate moved by 0.4%. Foreign exchange trading offices were in business which provided as few as 2306 crowns for $100 when 2511 crowns were available at a fully visible exchange

Table 1. Range of Crowns Available for a $100 Travelers Check in Prague Exchange Offices: September 11, 1996

	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*Total*
Between										
2300 and 2350	4	3	1	5	2	1	0	2	3	21
2351 and 2400	1	1	1	1	0	0	0	0	0	4
2401 and 2450	2	0	0	0	0	2	1	2	3	10
2451 and 2500	0	1	3	0	0	5	1	1	1	12
2501 and 2550	2	0	0	0	0	2	1	1	0	6
2551 and 2600	1	1	0	1	1	0	2	0	1	7
Total	10	6	5	7	3	10	5	6	8	60

Notes: A = Wenceslas Square; B = Old Town Square; C = Between Old Town Square and Charles Bridge; D = West of Charles Bridge; E = Mala Strana and the Castle Area; F = Main Train Station and Immediate Neighborhood; G = South of Wenceslas Square; H = Josefov; I = Between Wenceslas Square and Old Town Square.

trading office less than 100 yards away. Some tourists wasted 8.9% of their money in patronizing the former trading office.

This apparent violation of the assumption of rationality was investigated more fully with the help of a group of graduate students in the international finance course at the Center for Research and Graduate Education at Charles University during the Fall of 1996. Each of the students was assigned a neighborhood and was asked to survey up to ten foreign exchange trading offices during a 1 hour period on September 11, 1996. During that day, according to the *International Herald Tribune*, the interbank exchange rate moved from 26.35 crowns per dollar to 26.68 crowns per dollar. The students gathered information on the exchange rates which were offered for $100 dollar American Express travelers checks and on the commissions which were charged. They also noted whether the rates and commissions were easily seen from the street, whether the commission rates were shown in prominent or very small lettering, and the location of the offices.

Seventy five offices were surveyed, but 15 did not accept travelers checks and instead only exchanged paper currencies, leaving a total of 60 usable questionnaires. The results confirmed the similarity of exchange rates and the surprisingly wide range of commissions that were found in the earlier visits to a few offices on Wenceslas Square. In some cases very different commissions, and therefore numbers of crowns received for a $100 travelers check, were found in exchange offices literally next door to each other.

The most advantageous combination of exchange rate and commission among the 60 offices produced 2580 crowns for a $100 travelers check, while the least attractive outcome was 2306 crowns, a difference of 274 crowns or $10.27 at the interbank rate prevailing at the end of that day. The distribution of outcomes in each of the neighborhoods and for the entire sample can be found in Table 1.

The largest number of offices, many under the same brand names (Inter or Exact), offered the worst outcomes, but most of the neighborhoods had at least one office in the 'best' (over 2550 crowns) category. With regard to Wenceslas

Square, the office offering the best outcome (2563 crowns at Thomas Cook) was directly on the square at a prominent corner. It was no more than 50 yards from an office that managed to stay in business while offering 2330 crowns, a difference of 233 crowns or about $8.75 for a $100 travelers check. In both exchange offices the commission rates (zero on a rate of 25.63 for Thomas Cook and 9.75% on a rate of 25.82 for its neighbor) were prominently displayed.

The experience inside the main train station was equally interesting. The student who did that survey noted that three exchange offices were within 120 feet of each other along a hall leading from arriving trains, and that they offered combinations of exchange rates and commissions which produced as few as 2392 crowns and as many as 2530 crowns for a $100 travelers check. The difference, 138 crowns, represents $5.20. Some arriving tourists were wasting over 5% of their money in choosing the exchange office among the three neighbors with the least attractive combination of exchange rate and commission.

This student report was sufficiently surprising to justify an additional visit to the train station. Including those in the main hall, ten exchange offices were found offering a range from 2360 to 2530 crowns within the same building, which is 170 crowns or $6.40 on a $100 travelers check. On a return visit the author observed the original three offices with the 2392 to 2530 crown range when a passenger train was just arriving from Germany.

The behavior of those getting off the train was astonishing. Thirteen people undertook transactions at one of the three offices. Eight of them went to the office offering the worst outcome (2392 crowns), which happened to be the first one which a traveler would see when coming from the track on which the had train arrived. Two went to the next best office, which was 20 feet across the hall and which offered a saving of 29 crowns, or just over $1, on a $100 transaction. Only three people walked 120 feet toward an easily observable sign offering "low commission rates." If they had gone to the office next to that sign, they would have received 138 more crowns (about $5) than they received at the worst office. Not 1 of the 13 people checked the exchange rate and the commission at one office, and then went to another office. There was no comparison shopping.

How might an economist explain such apparently irrational behavior? Two possibilities come to mind; first, tourists may have tremendous confidence in the theory of perfect competition which they were taught in freshman economics; since many exchange offices exist in Prague, this must be a competitive market where prices are the same for all vendors, just as the principles of economics textbook said. Such confidence in the theory of perfect competition, however, should not survive more than one or two excessively expensive transactions; even tourists should learn from unhappy experience, and shift to cheaper exchange offices in sufficient numbers to force the expensive offices to lower their commissions. The fact that this does not happen leads to the second possibility.

Behaving in an economically rational manner requires mental effort, and people are inclined toward laziness. They simply do not want to be bothered, particularly when on vacation, with the need to think about saving money. Economists may enjoy such effort; normal people do not. Etzioni (1986) has argued that irrationality requires less effort than does rationality, and that therefore irrationality may be more natural for normal humans. This might be particularly likely for relatively small transactions, when people are on vacation and are therefore not interested in making serious mental effort to behave as an economist would expect. The data in this study would not lead one to predict how people would behave if they were buying a house or negotiating the compensation arrangements for a job. For relatively small transactions, however, particularly during vacations when people's primary goal is to enjoy themselves rather than work (think), behavior that economists view as irrational may be quite normal.

Acknowledgments: This paper was written while the author was a Visiting Professor at the Center for Research and Graduate Education at Charles University in Prague during the Fall of 1996. The author would like to thank Cristina Rata, Marek Fodor, Juliana Stadnikowa, Viliam Druska, Georgi Kambourov, Sergie Sontchik, Narine Nersesyan, Anna Loguinova, Daniala Tataru, and Pavel Loukiantsev for their help in gathering the survey information for this study.

NOTE

1. For discussions of altruistic economic behavior, see Frank, 1996; Manbridge, 1990. For articles questioning the assumption of rationality, see Stanley, 1994. See also Hanf, 1994; Etzioni, 1986; Devinney, 1889; and Salop and Stiglitz, 1977.

REFERENCES

Devinney, T. (1889). Rationally determined irrationality: An extension of the thesis of rationality as anti-entropic. *Journal of Economic Psychology, November*, 303-319.

Etzioni, A. (1986). Rationality in anti-entropic. *Journal of Economic Psychology, January*, 17-36.

Frank, R. (1996). What price the moral high ground? *Southern Economic Journal, July*, 1-17.

Hanf, C.H. (1994). Price, quantity, and consumers' behavior. *Journal of Consumer Policy, September*, 335-348.

Manbridge, J. (1990). *Beyond self interest*. Chicago: University of Chicago Press.

Salop, S., & Stiglitz, J. (1977). Bargains and ripoffs: A model of monopolistically competitive price dispersion. *Review of Economic Studies, 44*, 493-510.

Stanley, T.D. (1994). Silly bubbles and the insensitivity of rationality testing: An experimental illustration. *Journal of Economic Psychology, December*, 601-620.

56

WHARTON REPROGRAPHICS

chapter 14

THE BALANCE OF INTERNATIONAL PAYMENTS

During the course of a year the residents of one country engage in a vast number and variety of transactions with residents of other countries—exports and imports of merchandise and services, cash payments and receipts, gold flows, gifts, loans and investments, and other transactions. These transactions are interrelated in many ways, and together they comprise the international trade and payments of the national economy. Before we can analyze and evaluate a nation's international transactions, however, they must be classified and aggregated to make a balance of payments.

As a statistical classification and summary of all economic transactions between domestic and foreign residents over a stipulated period (ordinarily one year), the *balance of payments* of a nation affords an overall view of its international economic position. For this reason, the balance of payments is particularly helpful to government authorities—treasuries, central banks, and stabilization agencies—who are directly charged with the responsibility of maintaining external economic stability. Moreover, international trade is so important to many countries that the balance of payments must be carefully considered in the formulation of domestic economic policies, such as employment, wage, and investment policies.

The balance of payments of a country may also influence the decisions of business people. The experienced international trader or investor does not overlook the intimate bearing of the balance of payments upon the foreign exchange market and the course of government policy. A domestic exporter may hesitate to deal with an importer if it is suspected that the authorities of the importer's country will shortly impose or tighten exchange controls in the face of an adverse balance of payments. Dealers in foreign exchange also pay close attention to the balance of payments of countries whose currencies they handle in daily transactions. Failure to realize the close dependence of international business upon the balance of payments of the domestic and foreign countries has led to losses or even outright business failures.

THE COMPILATION OF THE BALANCE OF PAYMENTS

Three principles underlying the compilation of the balance of payments of a nation are worth special emphasis. First, only economic transactions between domestic and foreign *residents* are entered in the balance of payments. Second, a distinction is made between *debit* and *credit* transfers. Third, the balance of payments is a *double-entry* accounting statement.

The Concept of Residence

The balance of payments summarizes all economic transactions between domestic and foreign residents. Residence should not be confused with the legal notions of citizenship or nationality.

Individuals who represent their government in foreign countries, including members of the armed forces, are always considered residents of their own country. Thus, when a member of the American armed forces buys a glass of wine in France, an international transaction occurs that enters the balance of payments of both the United States and France.

Individuals who do not represent a government are considered to be residents of the country in which they have a permanent residence or in which they find their "center of interest." In some instances an individual's center of interest may be in doubt, but ordinarily such criteria as customary place of work, residence of employer, or principal source of income are sufficient to determine it. In the event of conflict, the permanent place of habitation takes precedence over center of interest. For example, an individual working at the United Nations in New York who does not represent a foreign government and who resides in New York permanently is treated as a resident of the United States despite the individual's foreign center of interest.

In preparing a balance of payments, the question of individual residence is much less important than the question of business residence. A corporation is a resident of the country in which it is incorporated, but its foreign branches and subsidiaries are viewed as foreign residents. Hence, shipments between an American concern and its overseas branch are international transactions and, as such, are entered in the U.S. balance of payments. At times the residence of a business may be difficult to decide. For example, a company may be incorporated in the domestic country, owned by residents of a second country, and conduct all of its business in a third country. In most instances, however, the residence of a business enterprise is readily apparent.

Government residence is the clearest of all: All government agencies are residents of their own country regardless of location.

Debit and Credit Entries in a Double-Entry System

With the exception of gifts, all international transactions represent a two-way exchange of assets between domestic and foreign residents.

Assets may be classified as *real* (goods and services) or *financial* (short- and long-term monetary claims, including money, which is the shortest-term claim). We can distinguish, therefore, three kinds of international transaction:

1. The exchange of goods or services for other goods or services. This is a *barter* transaction; both sides of the transaction are real.
2. The exchange of goods and services for money or other financial claims. These transactions are much more common than barter transactions. One side of the transaction is real and the other side is financial.
3. The exchange of financial claims, such as bonds or stock, for other financial claims, such as money. Both sides of this transaction are financial.

An international transaction, therefore, involves two opposing *transfers* of assets. Because the balance of payments is a double-entry accounting system, a transaction is entered twice: one transfer as a debit, the other transfer as a credit. *Hence, total debits and credits are always equal in a balance of payments.* This is an accounting equality that has no economic significance.

When is a transfer a debit entry? A credit entry? The accounting rules are depicted in Table 14-1.

A *debit* entry is made in a country's balance of payments for a transfer that either (1) *increases* the country's assets (something of value *owned*) or (2) *decreases* the country's liabilities (something of value *owed*). Conversely, a *credit* entry is made for a transfer that either (1) *decreases* the country's assets or (2) *increases* its liabilities.

A second way of distinguishing debit and credit transfers is as follows: A transfer that gives rise to money payments to foreign residents is a *debit* entry in the balance of payments; a transfer that gives rise to money receipts from foreign residents is a *credit* entry. This distinction, however, is less precise than the accounting distinction, and for that reason we shall use the latter in constructing the balance of payments.

Gifts are a special kind of international transaction and do not involve an exchange of assets. They are *unilateral transfers* of goods or services (gifts in kind) or of financial assets (money gifts). Despite this one-sidedness, gifts are entered twice in the balance of payments to preserve the double-entry accounting system. One entry records the actual unilateral transfer, and a second entry is made in a unilateral transfers account to record the missing side of the gift transaction.

Table 14-1 Accounting Rules for Entering Debits and Credits in the Balance of Payments

Direction of Change	*International Assets*	*International Liabilities*
Increase	Debit	Credit
Decrease	Credit	Debit

CONSTRUCTING THE BALANCE OF PAYMENTS

Table 14-2 depicts the conventional presentation of the balance of payments used by the International Monetary Fund (IMF). International transfers are grouped into three major accounts: the current account, the capital account, and the official reserves account. Entries in these accounts are recorded as *debits* or *credits*. The

Table 14-2 Conventional Presentation of a Balance of Payments (Dollars)

	Debits	*Credits*	*Net Debit (−) or Credit (+)*
A. Current account			
1. Goods	600	500	
		25	
Net balance on trade			− 75
2. Services			
(a) Transportation	100	50	
(b) Travel	75	60	
(c) Investment	50	100	
(d) Other	25	50	
Net balance on goods and services			− 65
3. Unilateral transfers	75	25	
Net balance on current account			− 115
B. Capital account			
1. Long-term capital			
(a) Direct investment	100	50	
(b) Portfolio investment	25	50	
2. Short-term capital	500	600	
	260	250	
	25	50	
	50	100	
	50	25	
		100	
Net balance on capital account			+ 215
C. Official reserves account			
1. Monetary gold (net)	5		
2. Other reserve assets (net)	100		
3. Liabilities to foreign central banks (net)		15	
Net balance on official reserves account			− 90
Net errors and omissions	10		− 10
Total debits and credits	2050	2050	

balance of payments is usually drawn up in the currency of the domestic country. Regardless of the currency in which they are made, international payments and receipts may be expressed in the domestic currency by the use of exchange rates for conversion from foreign to domestic currencies.

Current Account

All international transfers of *goods and services* are placed in the current account. Goods are physical products commonly referred to as merchandise. Services are intangible products, such as transportation, travel, capital services (as measured by interest and dividend income), construction, technical assistance, industrial property rights (as measured by income from patents, trademarks, and copyrights), and business services.

The current account also includes a *unilateral transfers* subaccount that records offsetting entries for real or financial unilateral transfers that appear elsewhere in the balance of payments. Unilateral transfers include institutional gifts for missionary and charitable purposes; personal gifts; and government transfers, such as grants to foreign countries to assist their economic development.

Table 14-2 shows three net balance entries in the current account: the net balance on trade (merchandise), the net balance on goods and services, and the net balance on current account.

Capital Account

The capital account includes all *financial* transfers between a country and the rest of the world, except for *official reserve transfers,* which are placed in a third major account.

One subaccount is *long-term capital*—financial claims and liabilities with an original maturity of more than one year. Long-term capital is further classified into direct investment and portfolio investment. *Direct investment* is defined as investment in enterprises located in one country but "effectively controlled" by residents in another country. Direct investment is a distinctive feature of international business because it usually involves the ownership of branches and subsidiaries abroad by domestic parent companies.[1]

Portfolio investment designates all long-term capital flows that do *not* give investors effective control over their investments.

Short-term capital—financial claims and liabilities, including money, with an original maturity of one year or less—includes international transfers of bank deposits and short-term securities, bank loans, and commercial claims and liabilities arising from the financing of current-account transfers.

---

[1] The U.S. Department of Commerce defines a direct investment as an investment that gives the investor at least 10 percent ownership of a foreign entity. However, this percentage is a judgmental statistical criterion, for there is no single ownership percentage that separates foreign investments that carry managerial control from those that do not (see Chapter 23).

Official Reserves Account

The official reserves account records the international financial transfers of the country's monetary authorities (the Treasury Department and the Federal Reserve System in the United States and central banks in other countries). The national monetary authorities hold the country's international reserves of gold and foreign exchange, which are used to provide residual financing of balance of payment transactions (known as "compensatory compensation") in a stable-rate system or to influence the exchange rate in a floating-rate system. These same authorities also issue currency and create the domestic reserves for commercial banks. Their responsibility for monetary policy (the cost and availability of credit) gives them a powerful role in external payment adjustment in addition to compensatory financing.

Official reserve transfers are classified as monetary gold, other reserve assets (mostly foreign exchange), and liabilities owing foreign central banks. The terminology used for monetary gold flows is identical to that used for merchandise. A monetary gold export is a credit; a monetary gold import, a debit. Since monetary gold is an international financial asset (a claim on foreigners), a gold export is equivalent to a capital inflow, while a gold import is a capital outflow. Ordinarily, however, gold movements are designated as such rather than as capital flows.

Errors and Omissions

Although the balance of payments is a double-entry account, governments do not draw up the balance of payments by recording each transaction twice. Governments simply do not have the necessary information on all of the many international transactions that take place over a year. Instead, they make single entries in the different accounts, depending on several sources of information that are not always consistent across accounts and vary in coverage and reliability. Consequently, *recorded* debits and credits are never equal in a country's balance of payments. To get that equality and thereby preserve the double-entry accounting system, an entry is made for "errors and omissions." In the balance of payments of the United States and other industrial countries, this entry is mainly attributable to inadequate coverage of service and short-term capital transfers. But in some developing countries, the major source of errors and omissions is the illegal smuggling of goods, such as the narcotics exports of Colombia.

ENTERING TRANSACTIONS IN THE BALANCE OF PAYMENTS

We can now demonstrate how transactions are entered in the balance of payments shown in Table 14-2. To facilitate our discussion, let us assume that this is the balance of payments of the United States, although the entries are illustrative only.

To make the proper entries, we need to answer five questions: (1) Is the transfer a *real* or a *financial transfer?* (2) If it is a real transfer, is it a *goods* or a *services* transfer? (3) If it is a financial transfer, is it an *official reserve* transfer or a *nonreserve* transfer? (4) If it is a nonreserve financial transfer, is it a transfer of *short-term* or *long-term* capital? (5) Is the transfer a *debit* or *credit* entry?

Entering Current-Account Transactions

We start with a U.S. merchandise export valued at \$500.[2] This export is a *real* transfer that goes into the current account (item A1). It is a *credit* entry because the export of merchandise *decreases* U.S. merchandise assets. (Alternatively, it is a credit entry because it gives rise to money receipts for U.S. residents.) The payment received from the foreign resident who bought the merchandise may take the form of dollars or of another country's currency. Both forms of payment are short-term *financial* transfers that are entered under item B2 of the capital account. The payment is a *debit* entry because it either decreases U.S. dollar liabilities owing foreigners (payment in dollars) or increases U.S. foreign-currency claims owed by foreigners (payment in a foreign currency).

A U.S. merchandise import of \$600 is recorded as a *debit* in the current account. Payment for this import by a U.S. resident is recorded as a *credit* in the short-term capital account. These entries are the converse of those for a merchandise export.

The purchase by U.S. residents of transportation services from foreign shipping firms (\$100), travel services abroad (\$75), the capital services of foreign-owned investment in the United States as measured by interest and dividends paid to foreigners (\$50), and other services from foreigners (\$25) are entered as *debits* in the current account. These are real transfers that increase the assets of U.S. residents. Payment for these service imports (amounting altogether to \$250) is entered in Table 14-2 as a single *credit* in the short-term capital account.

Conversely, the sale of services by U.S. residents to foreign residents is entered as a *credit* in the current account. Payments received by U.S. residents for these service exports (amounting to \$260) are entered in Table 14-2 as a single *debit* in the short-term capital account.

Unilateral transfers to foreigners over the past year, amounting to \$75, are entered as *debits* in the unilateral transfers account (A3). Keep in mind that this account is an accounting offset to the actual unilateral transfers, which take the form of goods and services or of money. Assuming that \$25 of unilateral transfers are in merchandise (for example, shipments of U.S. agricultural products under a U.S. government aid program) and \$50 in money, then \$25 is entered as a *credit* in the goods account and \$50 as a *credit* in the short-term capital account.

In the same year, unilateral transfers received by U.S. residents (\$25) are entered as a *credit* in the unilateral transfers account. Because we are assuming that the actual transfers consisted entirely of money, they are recorded in Table 14-2 as a single *debit* entry in the short-term capital account.

---

[2] A country's balance of payments is drawn up in its own currency. Hence, any transfers expressed in a foreign currency must be converted into the domestic currency at the current exchange rate.

We can now draw up the *net balances* in the current account. The net balance on trade is a *debit* of $75.[3] The net balance on goods and services is a *debit* of $65. The net balance on current account is a *debit* of $115.

Entering Capital Account Transactions

In addition to engaging in current-account transactions, U.S. residents during the year were lending to and borrowing from foreign residents. U.S. firms made *direct investments* totaling $100, as indicated by the *debit* entry in item B1(a). Direct investments abroad by U.S. firms are recorded as debits because U.S. residents *increased* their ownership claims, which constitute long-term capital assets. The firms financed their direct foreign investments in money (the most common form of payment), in goods and services, or a mix of both. We are assuming here that all payments to foreign residents for ownership claims were in money, and they are therefore recorded as *credits* in the short-term capital account.

Conversely, direct investment by foreign companies in the United States ($50) is recorded as a *credit* in item B1(a) of Table 14-2. The money received by U.S. residents from the sale of ownership rights to those foreign companies is recorded as a *debit* in the short-term capital account.

Over the year, U.S. residents also bought bonds and other long-term securities from foreign residents in the amount of $25. This *portfolio* investment is recorded as a *debit* in item B1(b) of Table 14-2 because it increases U.S. long-term claims on foreigners. Payment to foreign residents for these claims is recorded as a *credit* in the short-term capital account.

Conversely, foreign residents bought $50 of long-term securities from U.S. residents over the year. This inflow of portfolio investment is entered as a *credit* in item B1(b) because it increases U.S. long-term liabilities owing foreigners. Money receipts by U.S. residents from the sale of securities to foreigners are recorded as a $50 debit in the short-term capital account, representing an increase in U.S. short-term financial claims against foreigners.

So far all entries in the short-term capital account record payments or receipts for transfers entered in the current and long-term capital accounts. But short-term capital transfers also occur independently of those accounts. Banks, multinational enterprises, and persons located in the United States and abroad shift funds into and out of the United States to take advantage of interest-rate differentials among national money markets and to hedge against or speculate on movements in the exchange rate. Funds may also move out of countries judged to be high political risks into countries judged to be low political risks, a phenomenon called *capital flight*. Over the year, the United States received a *net inflow* of short-term capital, amounting to $100, due to these factors. This net inflow is entered as a *credit* in item B2 in Table 14-2 because it represents either a decrease in U.S. short-term claims on foreigners (as U.S. residents shift funds from abroad to the United

---

[3] A *debit* trade balance is commonly designated an *import balance* or a *trade deficit*. On the other hand, a *credit* trade balance is designated an *export balance* or a *trade surplus*.

States) or an increase in U.S. short-term liabilities owing foreigners (as foreign residents shift funds to the United States).

We can now calculate the net balance on capital account. Total debits add up to $1010; total credits, to $1225. The net balance, therefore, is a credit of $215.

Entering Official Reserves Transactions

Entries in the official reserves account are made on a net basis. Over the year, U.S. monetary authorities bought $5 of monetary gold from foreign central banks. This is regarded as a financial transfer because monetary gold is treated as a financial asset. Because an import of gold increases U.S. monetary gold assets, it is a *debit* entry. If the monetary authorities paid for this gold by crediting the accounts of the foreign central banks that sold it, then a *credit* entry is made in item C3 to record an increase in liabilities to foreign central banks.

The U.S. authorities also increased their holdings of other reserve assets (mostly foreign currencies) by a net of $100, as indicated by the *debit* entry in C2.[4] This increase in reserve assets was mostly offset by the net short-term capital inflow (credit entry) recorded in item B2. Lastly, U.S. liabilities owing foreign central banks increased by a net of $15, as shown by the *credit* entry in C3.

Entering Net Errors and Omissions

If we now add all the debit and credit entries in our balance of payments, we find that total credits exceed total debits by $10. Evidently, we have not recorded all transfers. Because the overall balance of payments *must* balance, we take care of this discrepancy by making a *debit* entry of $10 in the errors and omissions account. As noted earlier, the equality of total debits and credits has no economic significance.

Much of the errors and omissions entry derives from the comparatively poor record of private short-term capital movements. A radical shift in the size or sign of this entry from one payment period to the next is usually interpreted as a symptom of speculative or flight capital. Short-term capital may enter or leave a country outside the normal payment mechanism, especially when there is an intent to evade exchange controls and other official restrictions. In any event, short-term capital movements may either lead or lag behind the transactions they finance and thus show up in another payment period. For example, payment for a merchandise export shipped in December may not be recorded by a bank until January of the following year, or even later if the importer has received credit. Cash prepayment may cause the opposite situation.[5]

---

[4] In addition to foreign currencies (foreign exchange), other reserve assets include the U.S. reserve position with the IMF and special drawing rights (SDRs) issued by the IMF (see Chapters 18 and 19).

[5] Statistical discrepancies also appear in the balance of payments for the world as a whole. It is evident that the total of *all* countries' current-account balances should add up to zero because countries can only trade with each other. And yet there was a discrepancy of $45 billion in aggregate current-account balances in 1987. Portfolio investment income credits were $45 billion less than debits, and the gap between merchandise exports and imports was $36 billion. See "In a Maze of Numbers," *The Economist*, August 20, 1988, pp. 61–62.

BALANCE OF PAYMENTS ANALYSIS: EQUILIBRIUM AND DISEQUILIBRIUM

A balance of payments is always in accounting balance, but it is not always in *equilibrium*. The essential meaning of equilibrium is a sustainable relationship between two or more economic variables, such as the supply, demand, and price of an individual good. A balance of payments is in equilibrium when the nation's economy is in "fundamental adjustment" to the world economy. Conversely, when the balance of payments is in disequilibrium, the national economy must undergo changes in its price level, income, exchange rate, or other economic variables to restore a sustainable relationship with the rest of the world. The concept of balance of payments equilibrium is a market adjustment concept. Thus, official measures to control international transactions do not eliminate a disequilibrium but merely suppress it.

Three balances are most useful when analyzing a balance of payments: the balance on *current account*, the balance on *basic transactions*, and the balance on *official reserve transactions*. These balances are calculated by "drawing a line" at certain places in the balance of payments and then adding up all of the entries above the line. (The net balance on official reserves account equals the net balance on current and capital accounts but carries an opposite sign.) This is done in Table 14-3, which uses the same entries as appear in Table 14-2. To detect trends in

Table 14-3 Analytical Balances of the Balance of Payments (Dollars)

	Debits	*Credits*	*Net Balance*
A. Current account			
1. Goods	600	525	
2. Services	250	260	
3. Unilateral transfers	75	25	
Balance on current account			−115
B. Capital account			
1. Long-term capital			
(a) Direct investment	100	50	
(b) Portfolio investment	25	50	
Basic balance			−140
2. Short-term capital	910	1150	
3. Net errors and omissions	10		
Balance on current and capital accounts			+90
C. Official reserves account			
1. Monetary gold	5		
2. Other reserve assets	100		
3. Liabilities to foreign central banks		15	
Balance on official reserves account			−90

these balances, it is necessary to analyze a country's balance of payments over several periods, say, each year for three to five years. However, we shall confine our analysis to a single year.

Long-Run Equilibrium

Under what conditions are a nation's external economic relations sustainable over a period of time? If we take a very long time period measured in decades, the answer is plain. In the very long run, a nation *must* pay for its imports of goods and services through its exports of goods and services. That is to say, in the very long run a nation cannot depend on international credit (whether short- or long-term) to finance a net import balance on current account. Nor can it hope to sustain a net export balance. Long-run equilibrium demands, therefore, that the net balance on goods and services become zero over the secular time period.

Long-run disequilibrium is measured by the net balance on *current account.* A net current-account *deficit* (debit balance) indicates that a country is spending more than it is earning in transactions with other countries. Conversely, a current-account *surplus* (credit balance) indicates that a country is earning more than it is spending in its international transactions. Because a current-account deficit or surplus needs to be financed, the net balance on current account also measures the *net* international lending or borrowing of a country over the year. A deficit is financed by net borrowing from foreign residents; a surplus is financed by net lending to foreign residents. In Table 14-3, the net balance on current account is a deficit of $115, indicating that the United States is spending more than it is earning internationally and, therefore, is a net borrower in the same amount from the rest of the world.

Intermediate-Run Equilibrium

Ordinarily our time frame of reference in balance of payments analysis is much shorter than decades; it seldom exceeds three to five years, or about long enough to encompass short-run cyclical variations in national income. What are the conditions of balance of payments equilibrium over this time period? It is no longer necessary for the nation to balance imports against exports: It can sustain an import balance if *long-term* credits are available to finance it. Similarly, the nation can sustain an export balance on current account if it is willing to lend (or give) an equal amount on long term to the rest of the world. In such circumstances, there are no pressures for immediate adjustment of the balance of payments. Eventually, of course, a long-term borrower must repay the lender, but this is a future contingency that can be met or prepared for over a lengthy period of time.[6]

Intermediate-term disequilibrium in the balance of payments is measured by the net *basic* balance. The basic balance is the sum of the current account balance and the net movement of long-term capital. This balance is intended to measure the long-term or *structural* stability of a country's balance of payments. The transfers making up this balance are assumed to be insensitive to *short-run* changes in exchange rates and interest rates and to respond only to changes in factors of production, productivity, technology, consumer preferences, competition, and other real (nonmonetary) economic variables. A basic deficit (debit balance) or a basic surplus (credit balance) extending over several years are both indicators of an intermediate-run disequilibrium, because external payments cannot be sustained without adjustments in the economic relationship between a country and the rest of the world. Table 14-3 shows that the U.S. balance of payments is in fundamental disequilibrium with a basic deficit of $140.

---

[6] A family that finances a home on a twenty-five-year mortgage offers a crude analogy. As long as the family is able to service the mortgage, it is not forced to make adjustments in income, expenditures, or savings. The situation is sustainable.

Short-Run Equilibrium

In the short run (ranging from a moment to, say, a single year), equilibrium in the balance of payments exists when the monetary authorities make zero net use of international reserves: The amounts of foreign exchange in supply and demand are equal so there is no need for official intervention to sustain the exchange rate. Further, short-run equilibrium implies the absence of any government measures to adjust international payments through direct controls.

Short-run disequilibrium is measured by the net balance on *official reserves* account. As noted earlier, a country's monetary authorities use international reserves of gold and foreign exchange to maintain or influence the exchange rate. Therefore, the direction and size of the reserve transactions balance is an indicator of pressures on the exchange rate. A *credit* balance indicates that the authorities used their reserves to prevent or limit a *depreciation* of the country's currency. A *debit* balance reveals that the authorities added to reserves to prevent or limit an *appreciation* of the country's currency.

In Table 14-3, the basic deficit (a debit of $140) was more than offset by a net short-term capital inflow (including net errors and omissions) amounting to a $230 credit. The net balance on current and capital accounts, therefore, was a surplus (a credit of $90). This surplus was financed by a net *increase* of the same amount in the country's international reserves: Holdings of reserve assets (monetary gold and foreign exchange) received from foreign residents during the year rose by $105 while liabilities owing foreign central banks rose by only $15, making a *debit* net balance on official reserves account of $90.

Sources of Disequilibrium

To add further to our understanding of equilibrium and disequilibrium in the balance of payments, we offer brief descriptions of the sources of disequilibrium as classified in the following broad categories:

1. seasonal and random disequilibrium
2. cyclical disequilibrium
3. structural disequilibrium

4. destabilizing speculation and capital flight
5. other sources of disequilibrium

We shall discuss these categories as sources of *deficit* disequilibrium, but since a deficit disequilibrium in one country implies a surplus disequilibrium in one or more other countries, it should be clear that they are also sources of *surplus* disequilibrium.

Seasonal and Random Disequilibrium A nation's exports and imports vary seasonally due to seasonal changes in production and consumption, but usually this seasonal variation is not the same for both. The result is seasonal disequilibrium in the balance of payments. Seasonal disequilibrium is ordinarily of little consequence, since it is short-lived and self-reversible: A deficit in one season offsets a surplus in another.

Irregular, nonsystematic, short-lived disturbances may also cause disequilibrium in the balance of payments. The traditional example of this kind of disequilibrium is a crop failure that curtails exports or forces a nation to import foodstuffs. Labor strikes that tie up transportation or immobilize industries can also affect exports or imports and thus the balance of payments. Other sources of random disequilibrium include natural disasters such as floods and earthquakes. The foregoing random disturbances have a once-for-all impact on the balance of payments; they cause only a short-run disequilibrium. When disturbances are of such a magnitude as to have pervasive and lasting effects on the economy and trade of a nation (war, revolution, civil strife, and so on), then they are not random in the sense that we are using the term.

Cyclical Disequilibrium Variations in the national incomes of trading countries, whether involving changes in price levels, changes in the levels of production and employment, or both, may lead to cyclical disequilibrium in the balance of payments. In the early 1930s massive deflations in real income and production (accompanied by widespread unemployment) occurred in the industrial countries of North America and Western Europe. In the nonindustrial countries, where the agricultural sector accounted for the bulk of national income, deflation mainly took the form of a catastrophic fall in prices, wages, and terms of trade rather than in production and employment. This global deflation caused enormous payment disequilibria, wrecked the international monetary system, and forced nations into competitive depreciations and the wholesale use of trade and payment restrictions.

Since the end of World War II, the industrial countries have avoided profound deflation of the 1930s variety, but they have been far less successful in avoiding inflation generated by an effective demand running ahead of production. The situation in many developing countries has been far worse: "Runaway" inflation of 50 percent or more in a single year is not uncommon. Brazil and Argentina may be cited as examples. We can fairly say that inflation has been the single most important source of payment disequilibrium in the postwar years.

Cyclical disequilibrium may be a misnomer to denote payment disequilibrium caused by deflationary and inflationary movements in national income and price levels, because at most it is only roughly periodic. Usually, such movements are a source of intermediate-run disequilibrium lasting longer than a year.

Structural Disequilibrium Shifts in the demand for *specific* export goods, whether brought about by changing tastes, a new distribution of income in export markets, or changes in the availability or prices of competitive products offered by foreign suppliers, are common events in foreign trade. Shifts in the demand for import goods also occur for similar reasons. When a national economy is slow to adapt to these shifts, persistent structural disequilibrium appears in its balance of payments.

What is needed to remedy structural disequilibrium is a reallocation of production to conform to new patterns of demand and supply. Persistent structural disequilibrium may be viewed as a failure of the price system, which is supposed to guide and encourage reallocation in a market economy. Reallocation is sluggish when prices respond only slowly, if at all, to shifting market conditions because of monopoly, administrative pricing, price agreements, import protection, or other arrangements that limit price competition.

Adjustments in production usually demand a reallocation in the use of factors of production. In the export sector, for example, labor and other factors must move out of lines of production no longer in demand in foreign markets and into lines that enjoy a comparative advantage. When wages and the prices of other factors employed in declining export industries are hard to lower (sticky downwards), the reallocating function of factor prices is crippled and must depend on the attraction of higher factor rewards in growing industries in the export sector or elsewhere in the economy. A more serious deterrent to a prompt reallocation of factors is low factor mobility, especially labor. This creates the problem of structural unemployment.

A structural disequilibrium unprecedented in scope and severity struck the world economy when the OPEC countries quadrupled oil prices in the early months of 1974. The enormity of the payment burden suddenly imposed on the oil-importing countries was substantially responsible for the economic crisis of the mid-1970s. The further doubling of already high oil prices in 1979 undoubtedly contributed to the worldwide recession in the early 1980s. The magnitude and pervasiveness of the balance of payments adjustments to the high cost of OPEC oil places this structural disequilibrium in a class by itself.

To conclude, a wide variety of forces operating at home and abroad, such as technological and product innovations, new competition, changes in tastes, improvements in productivity, growing populations, and export cartels may cause shifts in the supply and demand of individual exports and imports that result in a structural disequilibrium in the balance of payments. Because it demands a reallocation of a country's productive capabilities, adjustment to structural disequilibrium is certain to be painful to displaced workers and industries, and therefore the restoration of international competitiveness usually takes several years. The continuing deficit disequilibrium in the U.S. balance of payments is partly structural in nature, reflecting the declining international competitiveness of the U.S. economy, as reviewed in Chapter 9.[7]

Destabilizing Speculation and Capital Flight Destabilizing speculation and capital flight (described in Chapter 12) constitute another source of payment disequilibrium. Although conceptually distinct, these short-term capital movements tend to occur together. Both intensify an existing disequilibrium, and capital flight may actually create one, since it is motivated by fears of safety that sometimes originate in conditions unrelated to the balance of payments, notably the prospect of war or revolution. In recent years the primary stimulus of destabilizing speculation and capital flight has been the expectation of depreciation in a major currency.

Other Sources of Disequilibrium The preceding four categories cover the main sources of disequilibrium, but they are not exhaustive. As we shall discover in Chapter 18, the U.S. payments deficit of the 1960s did not fit neatly into any one of these categories. And in the 1980s, an enormous current-account deficit in the U.S. balance of payments was financed by massive capital inflows induced by persistently high interest rates.

Disequilibrium may also originate in an unrealistic exchange rate, not uncommon when there is exchange control. Some economists also speak of *secular* disequilibrium, which arises from technological and other changes that occur slowly as an economy moves from one stage of growth to another over a period of decades.

Algebraic Summary of Equilibrium Conditions

In closing this section, we offer an algebraic summary of balance of payments equilibrium over the long run, the intermediate run, and the short run.

Over the long run, the equilibrium condition is $B_c = 0$, where B_c is the net balance on current account.

Over the intermediate run, the equilibrium condition is $B_c + B_{k,l} = 0$ where $B_{k,l}$ is the net balance on *long-term* capital account.

Over the short run, the equilibrium condition is $B_r = 0$, where B_r is the net balance on official reserves account.

THE NEED FOR BALANCE OF PAYMENTS ADJUSTMENT

A country's balance of payments cannot remain indefinitely in disequilibrium. But adjustment may take several forms. Over the short run, a country may adjust to an official reserves *deficit* by attracting inflows of short-term capital. But if the deficit persists, then a country must adjust in one of three ways: (1) through an internal deflation of prices and incomes relative to foreign prices and incomes, (2) through a depreciation of its rate of exchange, or (3) through an imposition of exchange and trade controls. Chapters 15, 16, and 17 investigate these three forms of adjustment to an international payments disequilibrium.

---

[7] The U.S. payments disequilibrium is also partly attributable to inappropriate macroeconomic policies. See Chapter 19.

Market versus Nonmarket Adjustment

The first two methods of adjustment work through market processes involving changes in income, prices, exchange rates, money supplies, interest rates, and other economic phenomena. Market adjustment, however, does not imply the absence of governmental action. Indeed, successful market adjustment to an external deficit depends upon government fiscal and monetary policies directed toward reinforcing equilibrating market forces. Otherwise government monetary and fiscal policies, while not controlling or supplanting market adjustment, may counteract its equilibrating effects on the balance of payments by introducing disequilibrating changes in income, prices, and the like.

Nonmarket adjustment stands in sharp contrast to market adjustment. Government controls and regulatory devices replace the market to suppress an external deficit. Controlled adjustment is not true adjustment. Although the symptoms of disequilibrium (net official reserves balance in the balance of payments or depreciation of the exchange rate) are removed by direct controls, the causes of disequilibrium (cyclical, structural, and so on) are left untouched. The result is *suppressed disequilibrium*. Suppressed disequilibrium generally results from the widespread use of import quotas and exchange restrictions. But restrictive measures and direct controls may take any number of particular forms: tied loans and grants, ceilings on direct investment outflows, generalized tariff surcharges, domestic buying preferences, and restrictions on the purchase of foreign bonds and on bank loans to foreigners, to name only a few.

Varieties of Market Adjustment

Market adjustment to persistent disequilibrium takes two different paths depending upon whether exchange rates are stable or free to vary. When exchange rates are held stable, either through adherence to the gold standard or through official stabilization agencies, adjustment occurs in the balance of payments. When, on the other hand, the exchange rate is free to respond to fluctuations in supply and demand, disequilibrium shows up as a one-way movement of the exchange rate. Adjustment occurs within the foreign exchange market (with repercussions on domestic and foreign income and prices), and the exchange rate moves until an equality is established between the amounts of foreign exchange supplied and demanded for current-account and long-term capital transfers.

When monetary authorities try to influence exchange rates but not to peg them at a particular level (a policy known as "dirty float" in contrast to a "pure float"), then adjustment occurs in both the balance of payments and the foreign exchange market. This is the situation today for the world's major currencies.[8]

---

[8] See Chapter 19.

SUMMARY

1. The balance of payments is a statistical classification and summary of all economic transactions between residents of one country and residents of other countries over a stipulated period of time, ordinarily one year. Residence is primarily determined by permanent location and secondarily by "center of interest."
2. International transfers are grouped into three major balance of payments accounts: the current account, the capital account, and the official reserves account. Entries in these accounts are recorded as debits and credits.
3. Because the balance of payments is a double-entry accounting statement, total debits always equal total credits. This accounting equality has no economic significance.
4. An international transfer that increases a country's assets or decreases its liabilities is a debit entry in the balance of payments. Conversely, an international transfer that decreases a country's assets or increases its liabilities is a credit entry.
5. With the exception of gifts, all international transactions are two-way exchanges of assets between domestic and foreign residents. Each side of a transaction is a transfer of real or financial assets.
6. All international transfers of goods and services (real transfers) are placed in the current account. Except for official reserve transfers, all financial transfers between a country and the rest of the world are placed in the capital account. Finally, the international reserve transfers of a country's monetary authorities are placed in the official reserves account.
7. The three balances most useful in the analysis of the balance of payments are the balance on current account, the balance on basic transactions, and the balance on official reserve transactions.
8. *Long-run* equilibrium demands a zero net balance on current account ($B_c = 0$); *intermediate-term* equilibrium requires a zero net basic balance ($B_c + B_{k,l} = 0$); *short-term* equilibrium occurs when there is a zero net balance on official reserves account ($B_r = 0$). Short-term equilibrium also means equilibrium in the foreign exchange market.
9. Sources of disequilibrium in the balance of payments may be seasonal and random, cyclical, structural, speculation and capital flight, and other.
10. When a country experiences a persistent deficit in its balance of payments, it cannot rely indefinitely on net short-term capital inflows and international reserves to finance that deficit. The need for adjustment policies that will eliminate the deficit becomes critical when a country's international reserves fall to a minimal level.
11. A government may use three different modes of adjustment to remove a persistent disequilibrium in its balance of payments, either singly or in some combination: adjustment through deflation, adjustment through devaluation, and adjustment through controls.

QUESTIONS AND APPLICATIONS

1. What is the balance of international payments?
2. What determines whether a transfer is entered as a debit or a credit in the balance of payments?
3. What is the distinction between current transactions, capital transactions, and official reserve transactions?
4. Using the double-entry system, record the following international transactions in a hypothetical U.S. balance of payments: (a) an export of U.S. merchandise, valued at $500, for payment in dollars by a foreign resident; (b) the purchase of hotel services, valued at $100, by an American tourist in France; (c) the sale by the U.S. monetary authorities of gold (worth $1,000) to the Bank of England; (d) the purchase of $200 of bonds by an American resident from a foreign resident; and (e) the acquisition by a U.S. multinational corporation of a firm in a foreign country in exchange for $1,000.
5. Why is the balance of payments always in an accounting balance?
6. What is the essential meaning of balance of payments equilibrium? Of disequilibrium?
7. Describe the major sources of disequilibrium.
8. What is the economic significance of the net balance on current account? The net balance on basic transactions? The net balance on official reserve transactions?
9. (a) Why is a country compelled to take adjustment measures to eliminate a persistent deficit in its balance of payments?
   (b) Which modes of adjustment may a country use to eliminate a persistent deficit?
10. Explain the difference between market and nonmarket adjustments.

SELECTED READINGS

International Monetary Fund. *Balance of Payments Manual.* Washington, D.C.: International Monetary Fund. Latest edition.

International Monetary Fund. *Balance of Payments Yearbook.* Washington, D.C.: International Monetary Fund. Annual.

Stern, Robert M., and others. *The Presentation of the U.S. Balance of Payments: A Symposium.* Essays in International Finance no. 123. Princeton, N.J.: Princeton University Press, 1977.

United States Department of Commerce, Office of Business Economics. *Survey of Current Business.* Washington, D.C.: U.S. Government Printing Office. March, June, September, and December issues, annual.

BIBLIOGRAPHY (continued)

Saunders, Anthony. *Financial Institutions Management: A Modern Perspective*, 2nd edition, Chapter 10, Boston: Irwin: Boston, 1997.

Senior Loan Officer Opinion Survey on Bank Lending Practices, Federal Reserve System, November 1996 and January 1997.

Talley, Karen. "Four-Lender Test Could Advance the Status of Credit Scoring," *American Banker* (March 24, 1997), p. 12.

Wantland, Robin. "Best Practices in Small Business Lending for Any Delivery System," *Journal of Lending and Credit Risk Management* (December 1996), p. 16-25.

Zuckerman, Sam. "Taking Small Business Competition Nationwide," *US Banker* (August 1996), pp. 24-28, 72.

What Determines the Exchange Rate: Economic Factors or Market Sentiment?

Gregory P. Hopper[*]

Readers of the financial press are familiar with the gyrations of the currency market. No matter which way currencies zig or zag, it seems there is always an analyst with a quotable, ready explanation. Either interest rates are rising faster than expected in some country, or the trade balance is up or down, or central banks are tightening or loosening their monetary policies. Whatever the explanations, the underlying belief is that exchange rates are affected by fundamental economic forces, such as money supplies, interest rates, real output levels, or the trade balance, which, if well forecasted, give the forecaster an advantage in predicting the exchange rate.

What is not so well known outside academia is that exchange rates don't seem to be affected by economic fundamentals in the short run. Being able to predict money supplies, central bank policies, or other supposed influences doesn't help forecast the exchange rate. Economists have found instead that the best forecast of the exchange rate, at least in the short run, is whatever it happens to be today.

---

[*] When this article was written, Greg Hopper was a senior economist in the Research Department of the Philadelphia Fed. He is now in the Credit Analytics Group at Morgan Stanley, Co., Inc., New York.

WHARTON REPROGRAPHICS

57.

In this article, we'll review exchange-rate economics, focusing on what is predictable and what isn't. We'll see that exchange rates seem to be influenced by market sentiment rather than by economic fundamentals, and we'll examine the practical implications of this fact. Sometimes, there are situations in which market participants may be able to forecast the direction but not the timing of the movement. We'll also see that volatility of exchange rates and correlations between exchange rates are predictable, and we'll examine the implications for currency option pricing, risk management, and portfolio selection.

THE EXCHANGE RATE AND ECONOMIC FUNDAMENTALS

The earliest model of the exchange rate, the monetary model, assumes that the current exchange rate is determined by current fundamental economic variables: money supplies and output levels of the countries. When the fundamentals are combined with market expectations of future exchange rates, the model yields the value of the current exchange rate. The monetary model might also be dubbed the "newspaper model." When analyzing movements in the exchange rate, journalists often use the results of the monetary model. Similarly, when Wall Street analysts are asked to justify their exchange-rate predictions, they will typically resort to some variant of the monetary model. This model is popular because it provides intuitive relationships between the economic fundamentals and it's based on standard macroeconomic reasoning.

The reasoning behind the monetary model is simple: the exchange rate is determined by the relative price levels of the two countries. If goods and services cost twice as much, on average, in U.S. dollars as they do in a foreign currency, $2 will fetch one unit of the foreign currency. That way, the same goods and services will cost the same whether they are bought in the U.S. or in the foreign country.[1]

But what determines the relative price levels of the two countries? The monetary model focuses on the demand and supply of money. If the money supply in the United States rises, but nothing else changes, the average level of prices in the United States will tend to rise. Since the price level in the foreign country remains fixed, more dollars will be needed to get one unit of foreign currency. Hence, the dollar price of the foreign currency will rise: the dollar will depreciate--it's worth less in terms of the foreign currency.

Money supplies are not the only economic fundamentals in the monetary model. The level of real output in each country matters as well because it affects the price level. For example, if the level of output in the United States rises, but other fundamental factors, such as the U.S. money supply, remain constant, the average level of prices in the United States will tend to fall, producing an appreciation in the dollar.[2] Future economic fundamentals also matter because they determine the market's expectations about the future exchange rate. Not surprisingly, market expectations of the future exchange rate matter for the current exchange rate. If the market expects the dollar price of the yen to become higher in the future than it is today, the dollar price of the yen will tend to be high today. But if the market expects the dollar price of the yen to be lower in the future than it is today, the dollar price of the yen will tend to be low today.

Here's an example of how to use the monetary model: suppose we wanted to predict the dollar-yen exchange rate. The first thing we need to do is think about the relationships between the fundamentals and the exchange rate. The monetary model implies that if the U.S. money supply is growing faster than the Japanese money supply, the dollar price of the yen will rise: the dollar will depreciate and the yen will appreciate. So, the analyst needs to assess monetary policy in the two countries. The monetary model also implies that if output is growing faster in the United States than it is in Japan, the dollar price of the yen will tend to fall: the dollar will appreciate and the yen will depreciate. Finally, the analyst must assess expectations about the future exchange rate. If the market's expectation of the future exchange rate were to change, the current exchange rate would move in the same direction. When making an exchange-rate forecast based on the monetary model, the analyst must consider the effect of all the fundamentals simultaneously. He can do this by using a statistical model or by combining judgment with the use of a statistical model.

In practice, using the monetary model to make exchange-rate forecasts is difficult because the analyst never knows the true value of the economic fundamentals. At any time, money supply and output levels are not known with certainty; they must be forecast based on the available economic data. Of course, expectations about the future of the exchange rate are even harder to assess because these expectations are unobservable. The analyst can always survey market participants about their expectations, but he can never be sure if the surveys accurately reflect the market's views. If we assume the monetary model is valid, the goal of the successful exchange-rate forecaster is to predict the values of the fundamentals better than the competition and then use the monetary model or some variant to derive forecasts of the exchange rate.

The fatal flaw in this strategy is the assumption that the monetary model can be used to successfully forecast the exchange rate once the values of the fundamentals are known. Although the monetary model had some early success, economists have established that the model fails empirically except perhaps in unusual periods such as hyperinflations.[3] For one thing, research did not establish a strong statistical relationship between exchange rates and the values of the fundamentals. Moreover, a key assumption of the model was found to be false: the model assumes that the price level can move freely. Yet the price level seems to be "sticky," meaning that it moves very slowly compared with the movement of the exchange rate.

What about other models? After the failure of the monetary model became apparent, economists went to work developing other ideas. Rudiger Dornbusch developed a variant of the monetary model called *the overshooting model,* in which the average level of prices is assumed to be fixed in the short run to reflect the real-world finding that many prices don't change frequently. The effect of this assumption is to cause the exchange rate to overshoot its long-run value as a result of a change in the fundamentals; eventually, however, the exchange rate returns to its long-run value. Ultimately, this model was shown to fail empirically: economists couldn't find the strong statistical relationships between the fundamentals and the exchange rate that should exist if the model were true.[4]

Another extension of the simple monetary model is called *the portfolio balance model.* In this approach, the supply of and demand for foreign and domestic bonds, along with the supply of and demand for foreign and domestic money, determine the exchange rate. Early tests of the model were not very encouraging.[5] Later, economists formulated a more sophisticated version of the portfolio balance model, in which investors were assumed to choose a portfolio of domestic and foreign bonds in an optimal way. According to the more sophisticated portfolio balance theory, the degree to which investors are willing to substitute domestic for foreign bonds depends on how much investors dislike risk, how volatile the returns on the bonds are, and the extent to which the returns on the different bonds in the portfolio move together. Unfortunately, economists did not find much empirical support for the more sophisticated version of the portfolio balance model.[6]

---

[1] When purchasing power parity holds, particular goods and services cost the same amount in the domestic country as they do in the foreign country. There is an extensive literature that documents that purchasing power parity doesn't hold except perhaps in the very long run.

[2] In the monetary model, the price level must fall in this situation to ensure that money demanded by consumers is the same as money supplied by the central bank.

[3] See the papers by Frenkel (1976, 1980), Bilson (1978), and Hodrick (1978) for empirical analysis of the monetary model.

[4] For an empirical treatment of the overshooting model, see the paper by Backus (1984).

Economic News. Thus, the three major models of the exchange rate—the monetary, the overshooting, and the portfolio balance models—do not provide a satisfactory account of the exchange rate. Nonetheless, it is possible that *news about the fundamentals* affects the exchange rate even if the fundamentals themselves don't influence the exchange rate in the manner suggested by the three major exchange rate models.

The news about the fundamentals can be defined as the difference between what market participants expect the fundamentals to be and what the fundamentals actually are once their values are announced. For example, market participants form expectations about the value of the money supply before the government announces the money supply figures, and these expectations are translated into decisions to buy or sell currency. These decisions ultimately help to determine the current level of the exchange rate. Once the government announces the value of the money supply, market participants buy or sell currencies as long as the news is different from what they expected. Thus, news about fundamentals, under this view, is an important determinant of the exchange rate.

The difficulty in testing this view is that economists don't know how to measure the news because they don't know how to measure the market's expectations. One solution is to assume that market participants form their expectations using a statistical device called linear regression. Using linear regression, an econometrician could estimate the expected level of a fundamental, such as the U.S. money supply, for each quarter during the past 20 years. He could then subtract the value of the estimated expected money supply from its actual value in each quarter to generate an estimate of the news about the quarterly U.S. money supply. The news for other fundamentals can be estimated in a similar way.

Once the econometrician has estimated each fundamental's news for each quarter during the last 20 years, he can check to see if it explains the level of the exchange rate. Studies by economists who have carried out this procedure generally indicate that news about the fundamentals explains the exchange rate better than the three major exchange-rate models.[7] However, two factors make this result hard to interpret. First, we have no direct evidence suggesting that market participants form their expectations using linear regression models or that they form their expectations as if they were using these models. Second, these studies use the final values of the fundamentals, values released by governments months, if not years, after the forecasts were made. Yet, forecasters must use the government's preliminary estimates of the fundamentals when they make their predictions. In other words, the econometrician is assuming that market participants are making forecasts using information they don't have. Hence, the result that news about the fundamentals seems to explain the level of the exchange rate better than the models is hard to interpret.

---

[5] See, for example, the paper by Branson, Halttunen, and Masson (1977).

[6] See the papers by Frankel (1982) and Lewis (1988) for empirical analysis of the more sophisticated portfolio balance model. The fundamental problem with the model is that investors must have an implausibly high aversion to risk to explain the exchange rate.

[7] For empirical analysis of news models, see the papers by Branson (1983), Edwards (1982, 1983), and MacDonald (1983).

One way to avoid the problem of using final values of fundamentals is to collect the initial estimates from newspapers, government announcements, and wire services and examine their ability to affect the level of the exchange rate. Studies that have done this have found that *announcements about fundamentals* affect the exchange rate only in the very short run: the effects of announcements generally disappear after a day or two.

When we look at the evidence from the three major exchange-rate models, from the news analysis, and from the effects of announcements, it is hard not to be pessimistic about the fundamentals' ability to explain the exchange rate. But the evidence we have examined so far is backward-looking: the fundamentals don't seem to explain exchange-rate behavior over the past couple of decades. However, we can also do a forward-looking analysis: do the fundamentals help us forecast the level of the exchange rate?

The surprising answer to this question, given by economists Richard Meese and Kenneth Rogoff in the early 1980s, is no. Meese and Rogoff examined the ability of the fundamentals to predict the level of the exchange rate for horizons up to one year. They considered fundamentals-based economic models as well as statistical models of the relationship between the fundamentals and the exchange rate that did not incorporate economic assumptions. They found that a *naive strategy* of using today's exchange rate as a forecast works at least as well as any of the economic or statistical models. Worse, they found that when they endowed the economic or statistical models with final values of the fundamentals—giving the models an advantage that forecasters could not possibly match—the naive strategy still won the forecasting contest. Despite many attempts since the publication of Meese and Rogoff's results, economists have not convincingly overturned their findings.

Thus, if we look backward or forward over periods of up to a year, the fundamentals don't seem to explain the exchange rate, contrary to what standard models in international finance textbooks imply. But this result might be dismissed by claiming that only the models tested have failed to explain the exchange rate. Perhaps economists will discover a model that works in the future.

Although a fundamentals-based model that works is a possibility, evidence from other countries suggests otherwise. In the European Exchange Rate Mechanism (ERM), exchange rates between major European currencies are kept relatively stable by the countries' central banks. If fundamentals are closely associated with the currencies, they should be stabilized as well. However, when we examine European fundamentals, we find that they fluctuate about as much as do the fundamentals of nonstabilized currencies, such as the U.S. dollar. Hence, the evidence from the European experience does not suggest a close connection between the fundamentals and the exchange rate, leading one to suspect that no fundamentals-based model will predict the short-run exchange rate.[8]

It's possible that the fundamentals really do explain the exchange rate, but we can't see the relationship because we can't observe the true fundamentals. Perhaps if economists discovered different economic models that use fundamentals other than money supplies and real output levels, the exchange rate could still be explained in terms of basic economic quantities. For example, some economic models imply that the true fundamentals are business technologies and tastes and preferences of consumers. However, the evidence from European countries renders this potential solution implausible. According to such a model, stabilization of European currencies in the ERM corresponds to stabilization of the true fundamentals. But why should business technologies and tastes and preferences of consumers change less in Europe than they do in the United States? At present, economists have found no evidence to suggest they do and, indeed, have little reason to suppose that they will ever find such evidence.

---

[8] See Rose (1994) for a detailed discussion of this point.

THE ALTERNATIVE VIEW: MARKET SENTIMENT MATTERS

The alternative view is that exchange rates are determined, at least in the short run (i.e., periods less than two years), by *market sentiment*. Under this view, the level of the exchange rate is the result of a self-fulfilling prophecy: participants in the foreign exchange market expect a currency to be at a certain level in the future; when they act on their expectations and buy or sell the currency, it ends up at the predicted level, confirming their expectations.

Even if exchange rates are determined by market sentiment in the short run, the fundamentals are still important, but not in the commonly supposed way. From reading the newspapers, we know that market participants take the fundamentals very seriously when forming exchange-rate expectations. Thus, if we wish to understand the level of the exchange rate, we need to know the values of the fundamentals and, more important, how market participants interpret those levels. However, the evidence we reviewed shows no pattern or necessary connection between the fundamentals and the level of the exchange rate. When market participants use the fundamentals to form expectations about the exchange rate, they don't use them in any consistent way that could be picked up by an economic or statistical model. As we have seen, we can do as well forecasting the exchange rate by quoting today's rate.

Although the naive forecast is at least as accurate as statistical or model-based forecasts, it's still not very good. It's just that statistical or model-based forecasts are so bad that even the naive forecast can do at least as well. How can we improve our forecast? Unfortunately, economists are just starting to build models of market sentiment, so we can't get much guidance from economic theory just yet. Nonetheless, we know that exchange rates are likely determined by market sentiment, so it seems reasonable to try to understand the psychology of the foreign exchange market to improve forecasts of the short-run exchange rate.

To understand the psychology of the foreign exchange market, we need to know about the various economic theories. Even if they aren't very accurate, their implications may still influence expectations in the market, although we would not expect any particular model to have any consistent influence. We also need to find out what the market is thinking. Probably the best way to do so is to be an active participant in the foreign exchange market and to talk to other participants to learn which events they think are important for a particular currency's outlook. These events might be announcements of fundamentals, political events, or some other factors. The analyst could then concentrate on forecasting those events. Of course, there will probably be no pattern to which events are important. For example, the U.S. budget deficit may well be important for the dollar one year and unimportant the next.

Speculative Attacks. In some cases, the forecaster might be able to make a reasonable guess about the direction of the exchange rate's movement, even if he can't be precise about the timing. As an example, let's review what happened to the exchange rate between the Swedish krona and the German deutsche mark in the early 1990s.

Sweden applied to enter the ERM in May 1991 in a bid to stabilize its currency. To stabilize the krona-deutsche mark exchange rate, interest rates in Sweden and Germany had to be the same. Therefore, the Swedish and German central banks couldn't independently use monetary policy—that is, change short-term interest rates—if they wanted to keep the exchange rate stable.[9] If Sweden wanted to act independently, it had to use fiscal policy (tax and government spending policies) to stimulate the country's growth rate.

However, a weak Swedish economy provoked speculators, who mounted an attack on the krona in September 1992. Speculators knew that the weak economy would tempt Sweden to abandon its fixed exchange rate and use monetary policy to cut short-term interest rates, especially since the new Swedish government was adopting restrictive fiscal policy. Speculators believed that if the Swedish central bank cut the short-term interest rate, the krona wouldn't be as attractive to investors. Thus, the speculators thought that after interest rates were cut, the currency would depreciate with respect to other ERM currencies. But since speculators expected the depreciation to happen, they decided to sell the currency immediately, i.e., mount a speculative attack on the currency.

This attack put the Swedish central bank in an uncomfortable position. To combat the currency's depreciation, the central bank raised short-term interest rates temporarily to repel the speculative attack—exactly the policy it didn't want in the face of sluggish economic growth. In fact, the Swedish central bank raised the short-term interest rate to an astonishing 500 percent and held it there for four days.[10]

The speculators were deterred, but not for long. The speculators understood that the Swedish central bank had to raise short-term interest rates temporarily to support the currency. But they were betting that the central bank wouldn't fight off the attack for long, especially in the face of disquiet in the country resulting from weak economic growth and the higher interest rates needed to fight the speculative attack. The high short-term interest rates had made the economic situation in Sweden even more precarious, so, in November, the speculators attacked again, selling the krona in favor of other ERM currencies. This time the Swedish central bank did not aggressively raise interest rates and the krona depreciated.

Profit opportunities such as this one can sometimes be exploited by speculators who recognize that a country's exchange-rate policy is inconsistent with the monetary policy needed, given a country's domestic situation. By paying careful attention to a country's economic and political developments, a speculator can sometimes forecast the direction of a currency's move when it breaks out of a stabilized exchange rate system. But the timing is not easily forecast; it is probably determined by market sentiment.[11]

---

[9] If a central bank can't change the short-term interest rate independently, it can't use monetary policy independently to stimulate the economy. Hence, countries with stabilized exchange rates must give up the independent use of monetary policy.

[10] If speculators expect the value of the currency to fall, and they are right, speculators can profit by selling the currency short. As an example, suppose a speculator anticipates that the value of the Swedish krona with respect to the deutsche mark will fall in one week. The speculator could borrow krona and sell them for deutsche marks at the current exchange rate. If the speculator is correct and the krona does depreciate, at the end of the week the speculator can buy back the krona for fewer deutsche marks than he sold them for. Provided the krona fell enough over the week, the speculator can repay the loan with interest and make a profit in deutsche marks. However, if the central bank makes short-term interest rates high enough, it can make this transaction unprofitable. Thus, one defense against a speculative attack is to dramatically raise short-term interest rates.

WHAT ABOUT TECHNICAL RULES?

Many market participants don't rely on the fundamentals. Instead, they use technical rules, which are procedures for identifying patterns in exchange rates. A simple technical rule involves looking at interest rates in two countries. Suppose the first country is the United States and the second is Canada. If the one-month U.S. interest rate is higher than the one-month rate in Canada, the U.S. dollar will tend to appreciate with respect to the Canadian dollar. But if the one-month Canadian interest rate is higher, the U.S. dollar will tend to depreciate with respect to the Canadian dollar. Economists and foreign exchange participants have often noted this fact.[12]

Indeed, it is possible to make money, on average, by using this rule. The problem is that implementing this rule carries risk. There is an ongoing debate about how big this risk is, and whether the average profits are explained by the level of risk. After all, it would not be surprising that the market pays a premium to those willing to assume substantial risk. Furthermore, the profits may have occurred only by chance and may not recur. Sometimes, economists report other technical rules that seem to make money in the foreign exchange market.[13] However, the considerations noted in the interest-rate differential rule apply to any technical rule. Even if the rule makes profits on average, the profits might be explained by the level of risk assumed in applying the rule. Moreover, the profits may well disappear when we account for technical statistical problems. Since economists are undecided at present about whether technical rules really do make money, it seems prudent to be cautious when evaluating the merits of any such rule.

WHAT ABOUT LONG-RUN FORECASTING?

Even though economic models or the fundamentals don't help us understand the exchange rate in the short run (except to the extent that they influence market psychology), there is evidence that models do better in the long run. For example, economists Martin Eichenbaum and Charles Evans report that currencies react as theory would suggest to unanticipated movements in the money supply, but only in the long run, after a period of about two years. Standard monetary theories would imply that an unanticipated decline in the U.S. money supply would lead to an appreciation of the dollar with respect to other currencies. Eichenbaum and Evans found that the dollar does, in fact, appreciate in response to an unanticipated monetary contraction; however, the full effects on the dollar are not registered until two years after the contraction, suggesting that models may well work in explaining the exchange rate in the long run.[14]

IS ANY ASPECT OF THE EXCHANGE RATE PREDICTABLE IN THE SHORT RUN?

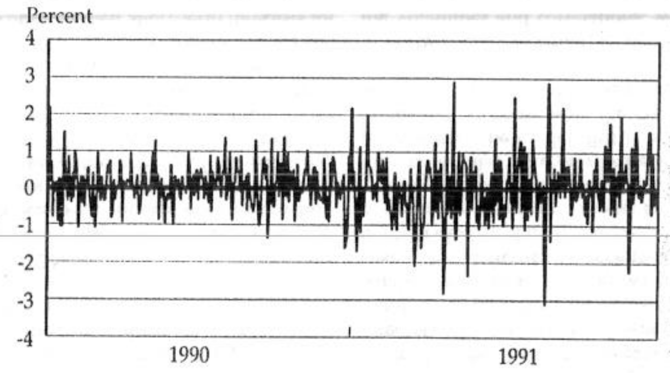
Although the level of the exchange rate in the short run is not very predictable, volatilities and correlations of currencies are much more predictable. The daily volatility of a currency measures the extent to which the currency's value in terms of another currency fluctuates each day. The value of high-volatility currencies fluctuates more each day than that of low-volatility currencies. Correlations measure the extent to which currencies move together. In general, volatilities and correlations vary with time, rising or falling each day in a somewhat predictable way.

The time-varying nature of the daily volatility of the dollar in terms of the deutsche mark can be seen in the figure. Notice that, in 1991, days on which the volatility of the dollar is high tend to cluster together, and in 1990, days with lower volatility follow one another. Since daily volatility clusters together, it is predictable. If we want to predict tomorrow's volatility, we need only look at the recent past. If daily volatility has been high over the recent past, we can be reasonably sure that it will be high tomorrow.

This idea forms the basis for statistical models of a currency's volatility. The GARCH model, developed by economist Tim Bollerslev, who built on work by economist Robert Engle, uses the volatility-clustering phenomenon to predict future volatility. In essence, a GARCH model measures the strength of the relationship between recent volatility and current volatility. Once this strength is known, it can be used to forecast volatility. GARCH models have good empirical support for exchange rates and are being used in practical applications in the foreign exchange market.[15]

GARCH models can be extended to handle two or more currencies, and they can measure the strength of recent correlations in predict-

Daily Percent Dollar Return on Deutsche Mark



---

[11] For further discussion of the myriad problems that can arise when countries attempt to fix their exchange rates, see the article by Obstfeld and Rogoff (1995).

[12] See my 1994 *Business Review* article for a nontechnical discussion.

[13] For an example, see Sweeney (1986).

[14] For further evidence on the effects of unanticipated monetary contractions on the exchange rate, see Schlagenhauf and Wrase (1995).

[15] GARCH stands for Generalized Autoregressive Conditional Heteroskedasticity. For the technical details of how GARCH models work, see Bollerslev (1986). Examples of technical applications of GARCH models of exchange rates include Bollerslev (1990) and Kroner and Sultan (1993). Heynen and Kat (1994) use GARCH to forecast volatility.

ing current ones. Once this strength is understood, it can be used to forecast correlations.

USES OF VOLATILITY AND CORRELATION FORECASTS

Volatility and correlation forecasts have important uses in finance. First, currency derivatives, securities whose value depends on the value of currencies, require measures of volatility and sometimes correlations to price them. GARCH models can supply estimates of these volatilities and correlations. Second, volatilities of individual currencies coupled with correlations between currencies can be combined to determine the volatility of a portfolio of currencies. Since the volatility of a portfolio measures the extent to which the portfolio's value fluctuates, the volatility can be used to assess a portfolio's risk. Portfolios with higher volatilities are riskier because they have a tendency to lose more per day—or gain more per day—than do portfolios with lower volatilities (see *Using GARCH to Measure Portfolio Risk*). Finally, knowledge of volatilities and correlations can help an investor choose the proportions of each currency to hold in a portfolio. For example, knowing a portfolio's volatilities and correlations may show an investor how to rearrange the proportions of currencies in a portfolio so that he has the same return, on average, but a lower risk of loss.

Using GARCH to Measure Portfolio Risk

Here, we illustrate the use of a GARCH model to manage risk in a simple portfolio of two currencies, the yen and the deutsche mark. Using daily data on the yen and the deutsche mark from January 2, 1981, to June 30, 1996, the time-varying volatilities and correlations were estimated using Engle and Lee's (1993a,b) GARCH model. Suppose we have a portfolio with $1 million invested in yen and $1 million invested in deutsche marks. Then we can calculate the value at risk (VaR) of the portfolio. The VaR is the maximum loss the portfolio will experience a certain fraction of the time during a specific period. For example, we can see from the table that daily VaR at the 95 percent confidence level is $12,000. That means that 95 percent of the time, the largest daily loss on the portfolio will be $12,000. But 5 percent of the time, the loss will be bigger, sometimes by a substantial amount. The daily loss measures the difference between the value of the portfolio at the end of one trading day and its value at the end of the next trading day.

As another example, consider weekly VaR at the 98 percent confidence interval. The numbers indicate that 98 percent of the time, the loss over five trading days will not exceed $35,000. But 2 percent of the time, the losses will be bigger. See Hopper (1996) for more discussion.

Value at Risk of a Currency Portfolio
with $1 Million Invested in Both Yen and Deutsche marks

	One-Day Horizon	Five-Day Horizon
95 percent	$12,000	$27,000
98 percent	$15,000	$35,000
99 percent	$18,000	$41,000

These numbers for the value at risk apply to the risk in the portfolio on July 1, 1996, the day after the end of the data period. However, the reason for using a GARCH model is that volatility varies over time. The value at risk would be higher in times of greater volatility and lower when the market is less volatile.

CONCLUSION

The evidence discussed in this article suggests that economic models and indeed fundamental economic quantities are not very useful in explaining the history of the exchange rate or in forecasting its value over the next year or so. This fact has important implications for market participants. It is all too common to encounter private-sector foreign exchange economists who tell very cogent stories designed to buttress their short-term forecasts for the values of currencies. These stories are often based on plausible economic assumptions or models. These economists hope that market participants will act on their forecasts and trade currencies. However, if these forecasts are justified by a belief that economic models or fundamentals influence the exchange rate in the short run, it's likely they are not very good. Indeed, we have seen that these forecasts will probably be outperformed by the naive forecast: tomorrow's exchange rate will be what it is today.

On the other hand, to the extent that these forecasts reflect market sentiment or a self-fulfilling prophecy, they may be useful. Unfortunately, it is difficult to judge when this is the case. The difficulty is accentuated by the unobservability of market expectations. A forecaster might be using a model he believes in, and his forecast might turn out to be correct if the market also temporarily believes the implications of the model. But it is hard, if not impossible, to know what the market expects; hence, it is hard to judge the merits of a forecast.

Fortunately, the situation is better regarding volatilities and correlations, which follow predictable patterns. The GARCH model and its more sophisticated variants can be used to price derivatives, assess currency portfolio risk, and set allocations of currencies in portfolios. Economists are continually discovering new empirical facts about volatility and correlations. No doubt the GARCH model will eventually be supplanted by an alternative, but for now, economists will use the GARCH model, or some variation of it, to forecast volatilities and correlations of currencies.

REFERENCES

Backus, David. "Empirical Models of the Exchange Rate: Separating the Wheat from the Chaff," *Canadian Journal of Economics* (1984), pp. 824-46.

Bilson, John. "The Monetary Approach to the Exchange Rate—Some Empirical Evidence," IMF *Staff Papers*, 25 (1978), pp. 48-75.

Bollerslev, Tim. "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31 (1986), pp. 307-27.

Bollerslev, Tim. "Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model," *Review of Economics and Statistics*, 72 (1990), pp. 498-505.

REFERENCES (continued)

Branson, William. "Macroeconomic Determinants of Real Exchange Rate Risks," in R.J. Herring, ed., *Managing Foreign Exchange Risk.* Cambridge, U.K.: Cambridge University Press, 1983.

Branson, William, Hannu Halttunen, and Paul Masson. "Exchange Rates in the Short Run: The Dollar-Deutschemark Rate," *European Economic Review,* 10 (1977), pp. 303-24.

Dornbusch, Rudiger. "Expectations and Exchange Rate Dynamics," *Journal of Political Economy,* 84 (1976), pp. 1161-76.

Edwards, Sebastian. "Exchange Rates and News: A Multi-Currency Approach," *Journal of International Money and Finance,* 1 (1982), pp. 211-24.

Edwards, Sebastian. "Floating Exchange Rates, Expectations, and New Information," *Journal of Monetary Economics,* 11, (1983), pp. 321-36.

Eichenbaum, Martin, and Charles Evans. "Some Empirical Evidence on the Effects of Monetary Policy Shocks on Exchange Rates," NBER Working Paper 4271 (1993).

Engle, Robert F. "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation," *Econometrica,* 50 (1982), pp. 987-1008.

Engle R., and G. Lee. "A Permanent and Transitory Component Model of Stock Return Volatility," Discussion Paper 92-44R, Department of Economics, University of California, San Diego (1993a).

Engle R., and G. Lee. "Long Run Volatility Forecasting for Individual Stocks in a One Factor Model," Discussion Paper 93-30, Department of Economics, University of California, San Diego (1993b).

Frankel, Jeffrey. "In Search of the Exchange Rate Risk Premium: A Six Currency Test Assuming Mean-Variance Optimization," *Journal of International Money and Finance,* 1 (1982), pp. 255-74.

Frenkel, Jacob. "A Monetary Approach to the Exchange Rate: Doctrinal Aspects and Empirical Evidence," *Scandinavian Journal of Economics,* 78 (1976), pp. 200-24.

Frenkel, Jacob. "Exchange Rates, Prices, and Money: Lessons From the 1920s," *American Economic Review,* 70 (1980), pp. 235-42.

Heynen, Ronald C., and Harry M. Kat. "Volatility Prediction: A Comparison of the Stochastic Volatility, GARCH (1,1), and EGARCH (1,1) Models," *The Journal of Derivatives,* 2 (1994), pp. 50-65.

Hodrick, Robert. "An Empirical Analysis of the Monetary Approach to the Exchange Rate," in J. Frenkel and H.G. Johnson, eds., *The Economics of Exchange Rates.* Reading, Mass.: Addison Wesley, 1978, pp. 97-116.

Hopper, Greg. "Is the Foreign Exchange Market Inefficient?" Federal Reserve Bank of Philadelphia *Business Review* (May/June 1994).

Hopper, Greg. "Value at Risk: A New Methodology For Measuring Portfolio Risk," Federal Reserve Bank of Philadelphia *Business Review* (July/August 1996).

Kroner, Kenneth F., and Jahangir Sultan, "Time-Varying Distributions and Dynamic Hedging with Foreign Currency Futures," *Journal of Financial and Quantitative Analysis,* 28 (1993), pp. 535-51.

Lewis, Karen. "Testing the Portfolio Balance Model: A Multi-lateral Approach," *Journal of International Economics,* 7 (1988), pp. 273-88.

MacDonald, Ronald. "Some Tests of the Rational Expectations Hypothesis in the Foreign Exchange Markets," *Scottish Journal of Political Economy,* 30 (1983), pp. 235-50.

Meese, Richard, and Kenneth Rogoff. "Empirical Exchange Rate Models of the 1970s: Do They Fit Out of Sample?" *Journal of International Economics,* 14 (1983), pp. 3-24.

Obstfeld, Maurice, and Kenneth Rogoff. "The Mirage of Fixed Exchange Rates," *Journal of Economic Perspectives,* 9 (1995), pp. 73-96.

Rose, Andrew. "Are Exchange Rates Macroeconomic Phenomena?" Federal Reserve Bank of San Francisco *Economic Review,* 1 (1994), pp. 19-30.

Schlagenhauf, Don, and Jeffrey Wrase. "Liquidity and Real Activity in a Simple Open Economy Model," *Journal of Monetary Economics,* 35 (1995), pp. 431-61.

Sweeney, Richard J. "Beating the Foreign Exchange Market," *Journal of Finance* (1986), pp. 163-82.

58

WHARTON REPROGRAPHICS

EXPLAINING EXCHANGE RATE BEHAVIOR: AN AUGMENTED VERSION OF THE MONETARY APPROACH

Thomas M. Humphrey

Prominent among competing theories of exchange rate determination in a regime of floating exchange rates is the monetary approach. This approach rests on the view that the exchange rate between two national currencies is determined by the respective national money supplies and demands in the two countries and the resulting effects on their general price levels.[1] To reach this conclusion the monetary approach combines the quantity theory of money with the purchasing power parity theory of exchange rates. The quantity theory says that the general price level is determined by the demand-adjusted money stock, i.e., by the nominal stock of money per unit of real money demand.[2] And the purchasing power parity doctrine holds that the exchange rate tends to equal the ratio of the price levels in the two countries concerned.[3] Taken together, the quantity theory and the purchasing power parity doctrine imply that the exchange rate is determined by relative demand-adjusted money stocks operating through relative national price levels.

The monetary approach has a long history dating back at least to the early 1800s when David Ricardo, John Wheatley, and other British bullionist writers used it to explain the fall of the paper pound on the foreign exchanges following Britain's switch from fixed to floating exchange rates during the Napoleonic wars. Later it was employed by the Swedish economist Gustav Cassel to explain the fall of the external value of the German mark during the famous hyperinflation episode of the early 1920s. Most recently, however, it has been employed, albeit with mixed results,[4] to explain the behavior of floating exchange rates in the post-1973 era of generalized floating.

The main shortcoming of the monetary approach is that it ignores the effect of real relative price changes on the exchange rate. In particular, it ignores the influence of changes in the *real terms of trade* (i.e., the relative price of imports and exports) and *internal relative prices* (i.e., the relative price of exports and domestic nontradeable goods), both of which function to clear national and international markets for real goods and services by equating commodity demand and supply. Determined by underlying shifts in consumer preferences, technology, and resource supplies, these real relative price changes necessitate equilibrium changes in exchange rates relative to the purchasing power parity ratio of nominal national price levels. Because the monetary approach assumes that purchasing power parity always holds, however, it cannot account for the influence of these real relative price changes on exchange rates. The result is that it ignores a key source of exchange rate disturbance, namely real shocks impinging on the exchange rate through the channels of the terms of trade and internal relative prices.[5] Far from recognizing these real channels, the monetary approach asserts that all factors affecting exchange rates must do so through monetary channels alone, i.e., through money supplies and demands. While this assumption may be warranted during periods of hyperinflation when exchange rate disturbances are of a predominantly monetary origin, it is clearly invalid in turbulent periods, such as the 1970s, when real shocks abound.

---

[1] For recent expositions of the monetary approach see Bilson [2], Frenkel [6], and Mussa [8].

[2] In other words, the price level equates money supply and demand by deflating the real value of the nominal money stock to the level people desire to hold.

[3] According to the purchasing power parity doctrine, this condition ensures that the common currency price of a standard basket of goods is everywhere the same so that there exists no arbitrage advantage to buying in one market over another. It also ensures that the real (exchange rate adjusted) purchasing power of both currencies is everywhere the same so that there exists no incentive to switch from one currency to the other and both moneys are therefore willingly held. The purchasing power parity doctrine argues that if these conditions were violated, goods would be cheaper in one country than another and one currency would be overvalued and the other undervalued on the foreign exchanges. The resulting rush to convert the former currency into the latter in order to purchase goods where they are relatively cheap would quickly bid the exchange rate back to the purchasing power parity at which no advantage exists to converting one money into the other.

[4] This is the conclusion reached by Kreinin and Officer [7: pp. 39-40] in their survey of empirical tests of the monetary approach. Of the 10 studies surveyed, at least 7 yield mixed or inconclusive results concerning the monetary approach. See also Stockman [10: pp. 675-6] who notes that the monetary approach has performed no better than simple purchasing power parity explanations and that "there remain substantial short-run variations in exchange rates unexplained by the monetary approach."

The foregoing shortcoming can be remedied by incorporating a real exchange rate component into the analytical framework of the monetary approach. The result is an augmented monetary model that captures all factors, real and monetary, affecting exchange rates. This article constructs such a model, discusses its constituent components, and uses it to explain certain characteristic features and policy implications of observed exchange rate behavior in recent years.

The Model and Its Components The model itself assumes two hypothetical open national economies each with its own currency and each producing two goods, namely (1) a purely domestic (nontradeable) good and (2) a unique exportable good, part of which is consumed domestically and part of which is exported to the other country. The basic building blocks of the model include (1) a terms of trade identity that links the exchange rate with export prices, (2) a price structure identity that links export prices to general prices via a term containing the relative price of exportables and nontradeables, (3) a quantity theory equation that links general prices to money supply and demand, and finally (4) a money demand equation that links the demands for foreign and domestic currencies to the expected rate of change of the exchange rate. Taken together, these components imply that the exchange rate is determined by the multiplicative product of the terms of trade, relative price structures, relative nominal money stocks, and relative real money demands, respectively. Of these four determinants, the first two constitute the so-called real or price-deflated exchange rate that captures the effect of real disturbances operating through nonmonetary channels. By contrast, the last two determinants constitute the nominal or monetary element of the exchange rate. As such they capture the effect of monetary, real, and expectational disturbances operating through monetary channels, i.e., through money supply and demand.

The foregoing variables are denoted by the following symbols: let E be the observed market exchange rate (defined as the domestic currency price of a unit of foreign currency), R the real or price-deflated value of that exchange rate (i.e., the exchange rate divided by the purchasing power parity ratio of national price levels), and e the expected future rate of change of the exchange rate. Furthermore, let M be the nominal money stock (assumed to be exogenously determined by the central bank) and D the real demand for money, i.e., the stock of real (price-deflated) cash balances that the public desires to hold. Also let T be the real terms of trade (defined as the quantity of exports given up per unit of imports obtained) and S be the structure of prices in each country as represented by the relative price of exportable goods in terms of the general price level. Finally, let P_x be the price of exportable goods, P_n the price of nontradeable goods, and P the general price level defined as the weighted average of the prices of exportable goods and nontradeable goods, respectively (i.e., the aggregate price of gross domestic product). Asterisks distinguish foreign-country variables from home-country variables. The foregoing variables are linked together via the model's basic building blocks described below.

Terms of Trade The first building block of the model is the concept of the real terms of trade. Representing the quantity of exports that must be sacrificed to obtain a unit of imports, the terms of trade T is defined as the relative price of imports and exports (i.e., the ratio of import prices to export prices). Since the domestic currency price of goods imported from abroad is the same as their foreign currency export price multiplied by the market exchange rate between domestic and foreign currency, the terms of trade may be defined as

$$(1)\quad T = E\, P_x^*/P_x,$$

where T is the terms of trade, E the exchange rate, P_x^* the foreign currency price of foreign country exportables, and P_x the domestic currency price of domestic exportables. Via this identity the terms of trade variable links the exchange rate to export prices in both countries as can be seen by rewriting the identity as $E = T\, P_x/P_x^*$.

---

[5] Note, however, that the monetary approach does capture the effect of real shocks operating through the real income determinant of the demand for money. That is, the monetary approach captures the income effects but not the relative price effects of real disturbances.

Being a real relative price, the terms of trade is affected by real economic variables such as productivity, consumer preferences, resource supplies, and the structure of particular markets in both countries. For example, if productivity is increasing faster in A's export-producing sector than in B's, the consequent rise in A's export supply relative to B's means that A must give up more exports per unit of imports obtained from B. As a result, A's export price will fall relative to B's and the terms of trade will turn against A. Likewise a shift in world demand from A's exports to B's exports will raise the relative price of the latter and worsen A's terms of trade. Finally, suppose B's export sector becomes monopolized while A's remains competitive. B's exporters could exploit their newly acquired market power by restricting output and raising prices, thereby turning the terms of trade against A. In general, the greater the relative productivity and the lower the degree of market power in A's export sector compared with B's, and the lower the intensity of demand for A's exportables relative to B's, the worse the terms of trade for A and vice versa for B.

Internal Price Structure The second building block of the model is the concept of the internal price structure that links export prices to general prices via a term that summarizes the composition of relative prices in each country. Written as follows

$$(2) \quad P_x = SP \text{ and } P^* = S^*P_x^*$$

the internal price structure S is the ratio of export prices to general prices as can be seen by rewriting the identities as $S = P_x/P$ and $S^* = P_x^*/P^*$. Defined as the relative price of exportables in terms of the general price level, the internal price structure S is also the relative price of exports and nontradeable goods. To show this, write the general price level as a weighted geometrical average of the respective prices of the two goods, i.e.,

$$(3) \quad P = P_n^a P_x^{1-a}$$

where P denotes general prices, P_n nontraded goods prices, P_x exportable goods prices, and the weights a and 1—a denote the shares of the two goods in the gross domestic product. Dividing both sides of this expression by P_x and inverting the result yields

$$(4) \quad P_x/P = (P_x/P_n)^a = S$$

which says that the internal price structure S is equivalent to the relative price of exportable and nontradeable goods weighted by the latter's share in the gross domestic product.

Stated this way, the price structure variable measures the internal opportunity cost of producing exportables such that a rise in S means that a country will have to give up more nontradeables per additional unit of exportables produced. And, when combined with the terms of trade, it also measures the opportunity cost of transforming nontradeables into imports by way of exportables. Thus if Q_n/Q_x represents the quantity of nontradeables given up to produce a unit of exportables (the internal price structure) and Q_x/Q_m is the quantity of exportables sacrificed to obtain a unit of imports (the terms of trade), it follows that the product of these ratios $(Q_n/Q_x)\ (Q_x/Q_m) = Q_n/Q_m$ shows the domestic nontradeables cost of obtaining imports by means of exports. In this regard the price structure variable represents the *indirect* terms of trade just as the relative price of exports and imports represents the *direct* terms of trade.

Being a real relative price, the internal price structure is affected by real economic variables such as intersectoral differences in productivity, tastes, and the degree of market power. For example, if productivity (and hence output) is advancing faster in a country's export-producing sector than in the rest of the economy, the resulting rise in the relative supply of exportables will lower their internal relative price thereby altering the price structure. Likewise a reduction in the degree of market power in the export sector relative to the rest of the economy will result in a fall in the relative price of exportables and a corresponding change in the internal price structure. Similarly, a shift in demand away from a country's exportable good to its nontradeable good will lower the internal relative price of exportables and alter the structure of prices.

Before proceeding to the third building block of the model, it should be noted that substituting equation 2 into equation 1 and solving for the exchange rate yields

$$(5) \quad E = T\frac{S}{S^*}\frac{P}{P^*}$$

which expresses the exchange rate as the product of the terms of trade, relative price structures, and the purchasing power parity ratio of national price levels, respectively. Regarding this expression three points should be made. First, it recognizes that factors other than national price levels affect exchange rates. In particular, it says that the purchasing power parity ratio of national price levels is *a* determinant but not *the* sole determinant of exchange rates. In this respect it differs from the simple monetary approach, which identifies the exchange rate with the purchasing power parity determinant alone.

Second, equation 5 specifies the terms of trade and relative price structures as the constituent components of the real (price-deflated) exchange rate. That is, since the real exchange rate is by definition the observed exchange rate divided by the purchasing power parity ratio of national price levels, it follows that the real exchange rate R is the product of the terms of trade and relative internal price structures as can be seen by writing the equation as $E/(P/P^*) = R = TS/S^*$. This real exchange rate changes with shifts in the terms of trade and internal price structures. It also undergoes temporary changes when sluggish price adjustment prevents national price levels from responding as fast as exchange rates to changes in underlying economic conditions.

Third, the equation shows that the strict purchasing power parity condition assumed by the monetary approach holds only if the real exchange rate R is unity. This can be seen by rewriting equation 5 as

$$(6) \quad E = R\frac{P}{P^*}$$

which says that the exchange rate will equal the purchasing power parity ratio of national price levels only when the real exchange rate is one. But the real exchange rate will be unity only in the special case in which both countries produce a single identical traded good (or standard basket of traded goods) such that commodity arbitrage will render the real price of this good everywhere the same. In all other cases the real exchange rate can be expected to possess a value other than unity. Hence we conclude that the strict purchasing power parity condition postulated by the monetary approach rests on the special assumption of a one-good world.[6]

Quantity Theory of Money The third building block of the model is the quantity theory of money which links general prices P to national nominal money supplies M and real money demands D. Written as follows

$$(7) \quad P = M/D \text{ and } P^* = M^*/D^*$$

the quantity theory says that the general price level in each country is determined by the demand-adjusted money stock, i.e., by the nominal stock of money per unit of real money demand. Written in the form $M/P = D$, the quantity theory also expresses the condition of money market equilibrium according to which the price level adjusts to equate the real (price-deflated) value of the nominal money stock with the public's real demand for it thereby clearing the market for real cash balances. Note that equation 7 implies that the purchasing power parity ratio of domestic to foreign national price levels is determined by relative national money supplies and demands. Here is the essence of the simple monetary approach to exchange rate determination, namely the extension of the quantity theory of money to the open economy under floating exchange rates.

Money Demand Functions The fourth component of the model consists of money demand functions linking the demands for foreign and domestic currency to the expected future rate of change of the exchange rate. Money demand is assumed to be a function of four variables, including (1) real income (a proxy for the transaction demand for money), (2) nominal interest rates (the opportunity cost of holding money rather than bonds), (3) expected future rate of inflation (the anticipated depreciation cost of holding money rather than goods), and (4) expected rate of change of the exchange rate (the anticipated rate of return from holding foreign money rather than domestic money). In what follows, however, all but the last of these money demand determinants are suppressed and real money demands D in both countries are treated as a function solely of the anticipated future rate of change e of the exchange rate, i.e.,

$$(8) \quad D = D(e) \text{ and } D^* = D^*(e).$$[7]

Equation 8 emphasizes the crucial role of exchange rate expectations in the determination of current exchange rates. It implies that exchange rates behave as efficient asset prices, being extremely sensitive to expectations of future conditions and adjusting instantaneously to changes in those expectations. In particular, it states that money demand functions provide the channel through which expectations influence exchange rates, i.e., expectations determine relative demands for the two currencies and therefore also the exchange rate between them. Thus a rise (fall) in the expected future rate of change of the exchange rate will, by raising (lowering) the expected yield from holding foreign rather than domestic currency, shift demand to the former (latter) thereby depreciating (appreciating) the current exchange rate. In this way current exchange rates are determined by exchange rate expectations operating through the channel of relative national money demands.

---

[6] See Sakakibara [9; p. 204] for a discussion of this point.

[7] Equation 8 enters exchange rate expectations directly into the money demand function on the grounds that such expectations constitute the anticipated depreciation cost of holding one currency over the other. The same result can be derived indirectly by assuming (1) that the demand for each currency is determined by the nominal interest rate on securities denominated in that currency, (2) that international nominal interest rate differentials equal the forward premium on foreign exchange (the interest rate parity condition), and (3) that the forward premium equals the expected rate of depreciation of the exchange rate. This latter interpretation views the expected rate of change of the exchange rate not as the cost of holding one currency over the other but rather as the relative opportunity cost of holding either currency instead of securities. Both interpretations yield the same conclusion, namely that expectations of future exchange rates influence current exchange rates through real money demands.

Components of the Exchange Rate The foregoing elements can be combined into a single reduced form expression by substituting equations 2, 7, and 8 into equation 1 and solving for the exchange rate. The resulting expression is

$$E = T\frac{S}{S^*}\frac{M}{M^*}\frac{D^*(e)}{D(e)} \tag{9}$$

which says that the exchange rate is composed of the product of four groups of factors, namely the real terms of trade, relative internal price structures, relative nominal money supplies, and relative real money demands, respectively. Of these four components, the first two comprise the real exchange rate and the last two the underlying determinants of the nominal or purchasing power parity exchange rate. More precisely, the terms of trade and relative price structure variables account for real influences affecting the exchange rate through nonmonetary channels. By contrast, the relative nominal money stock variable accounts for purely monetary influences affecting the exchange rate through monetary channels (i.e., through money supplies). Finally, relative real money demands account for real and expectational influences affecting the exchange rate through monetary channels (i.e., through money demands).

Regarding the effect of these four determinants on the exchange rate, the equation predicts that a rise in each will tend to depreciate the exchange rate and a fall to appreciate it. That is, the equation predicts that a country's currency will depreciate upon:

(1) a worsening of the terms of trade (i.e., a rise in the export cost of obtaining imports),

(2) a rise in the relative price structure reflecting an increase in the nontradeables cost of producing exportables,

(3) a rise in the relative money stock due to a faster rate of monetary expansion at home than abroad, and

(4) a rise in the demand for foreign relative to domestic money due, say, to a rise in the expected future rate of depreciation of the exchange rate.

Conversely, the equation indicates that the exchange rate will appreciate given (1) an improvement in the real terms of trade, (2) a fall in the nontradeables cost of producing exportables, (3) a reduction in domestic relative to foreign money growth, and (4) a rise in domestic relative to foreign money demand reflecting improved prospects for the value of the domestic currency.

Application of the Model Having outlined the augmented monetary model, the next step is to use it to answer certain questions arising from exchange rate experience in the post-1973 era of floating exchange rates. The first question is: What has caused the large exchange rate fluctuations observed in recent years?[8]

The model outlined above identifies three sources of exchange rate disturbance, namely

- real shocks operating through the terms of trade and relative internal price structures,
- monetary and real shocks operating through money supplies and demands, and
- changes in exchange rate expectations operating through relative money demands.

All three types of shocks were prevalent in the turbulent 1970s and all three contributed to exchange rate movements. Real shocks occurred in the form of oil embargoes, changes in international demands, commodity shortages, tax and regulatory burdens, shifts in commercial policy, productivity growth differentials and the like. Monetary shocks occurred in the form of divergent money growth rates and frequent sharp shifts in short-term policy targets. Also, during this period uncertainty about future developments became more intense. The policy surprises and the associated increased uncertainty about the future induced large and frequent changes in exchange rate expectations. Channeled through real money demands, these expectational changes were immediately embodied in the price of foreign exchange which jumped to its new equilibrium level consistent with the altered expectations. In short, the events of the 1970s indicate the extent to which disturbances can affect exchange rates. Given the abundance of shocks, surprises, and uncertainties in the post-Bretton

---

[8] See Artus and Young [1; pp. 23-33] for a discussion of these fluctuations.

Woods period it is small wonder that exchange rates moved as much as they did.

Exchange Rate Volatility The second question refers to the high degree of short-run (daily, weekly, monthly) volatility exhibited by exchange rates. With respect to volatility, Frenkel [5; p. 23] and Flood [3; pp. 10-13] note that since the adoption of floating exchange rates in early 1973 exchange rates have displayed a degree of variability far exceeding that of national price levels and sometimes approaching that of stock prices quoted on the securities exchanges. And Artus and Young [1; p. 26] argue that exchange rates have been much more volatile than their underlying economic and financial determinants (e.g., money growth rate differentials, real relative prices, inflation differentials and the like). Why are exchange rates so volatile?

Exchange rates are more volatile than their basic economic determinants because they depend not only upon the current value of those determinants themselves but also upon expectations of the entire future paths of those determinants. This expectational factor magnifies the impact of unanticipated changes in economic conditions on the exchange rate. Those changes affect the exchange rate directly and also indirectly through their effect on expectations. In particular, by inducing shifts in exchange rate expectations, disturbances to underlying economic conditions may engender large and frequent movements in exchange rates. For example, unforeseen changes in monetary growth rates may, by altering expectations of future monetary conditions, produce disproportionally large changes in current exchange rates. Viewed this way, exchange rate volatility is seen to stem from large and frequent shifts in exchange rate expectations operating through relative money demands.

In this connection, the model stresses that exchange rates are efficient asset prices dominated by expectations of the future and extremely sensitive to new information (announcements, rumors, unforeseen events, policy surprises and the like) that alters those expectations. Consequently, when new information appears it is immediately discounted into the market price of foreign exchange which jumps to its new equilibrium level consistent with the changed expectations. Since new information abounds by definition in periods of turbulence and uncertainty, it follows that exchange rates will exhibit a large degree of volatility in such periods. The 1970s constitute a prime example of such a period. Given the economic shocks, political upheavals, policy surprises, and uncertainty that dominated that period, the observed volatility of exchange rates is no mystery. That volatility reflected the large and frequent shifts in expectations induced by the shocks, surprises, and news of that period.

Sources of Departure From Purchasing Power Parity The third question is: Why have exchange rates since 1973 failed to conform to the predictions of the purchasing power parity theory? That theory predicts that exchange rates will move over time as the ratio of nominal national price levels so as to leave the real value of money and its counterpart, the real price competitiveness of goods, everywhere the same. These predictions have not held up well in recent years. On the contrary, as Jacob Frenkel convincingly demonstrates in his paper "The Collapse of Purchasing Power Parities During the 1970s" [4], exchange rates have frequently deviated sharply from purchasing power parities and in many cases these deviations have persisted with the passage of time. What caused these deviations that have produced such large discrepancies between the external and internal values of currencies?

The model suggests that deviations from purchasing power parity occur for two reasons. One is nonsynchronous movements of exchange rates and price levels due to sluggish price adjustment; exchange rates adjust much faster than national price levels to changes in underlying economic conditions. Being efficient asset prices, exchange rates are extremely sensitive to unforeseen changes that alter expectations of the future. By contrast, national general prices are composed largely of commodity prices reflecting past and present conditions as embodied in existing contracts and are therefore relatively unresponsive to unforeseen changes in economic conditions. Consequently, when changes occur, sensitive exchange rates adjust immediately whereas sluggish national price levels lag behind. The resulting differential speed of price response causes a temporary divergence from purchasing power parity. It follows that in turbulent periods like the 1970s, when shocks and surprises occur frequently, exchange rates will deviate from purchasing power parity much of the time.

Sluggish price adjustment is not the only source of deviation from purchasing power parity, however. The model suggests that real structural changes in tastes, technology, and market structure also play a role. Operating through real relative prices, these structural changes necessitate real equilibrium changes in the exchange rate and thereby produce systematic divergences from purchasing power parity. In terms of the exchange rate equation $E = RP/P^*$, real structural changes generate movements in the real exchange rate R, thereby necessitating equilibrium shifts in the market exchange rate E relative to the purchasing power parity ratio of national price levels.

To illustrate how real structural changes produce systematic deviations from purchasing power parity, consider a hypothetical case in which the United States exports wheat to OPEC in exchange for oil. For convenience, assume that the monetary authorities are stabilizing general prices in both countries so that no exchange rate disturbances arise from that source. That is, all shifts in money demand (due, for example, to productivity-induced increases in income) are accompanied by corresponding changes in the money supply so as to leave general prices and purchasing power parity unchanged. Now suppose that OPEC forms a cartel and quadruples the price of oil. The rise in the price of oil relative to the price of wheat means that the United States now has to export more wheat than before to obtain a barrel of imported oil. The resulting worsening of the U. S. terms of trade and the consequent rise in the real exchange rate causes the dollar to depreciate relative to OPEC currencies despite no underlying change in purchasing power parity. In short, a real shock in the form of increased market power induces a real exchange rate depreciation relative to the unchanged purchasing power parity.

As another example, suppose that U. S. demand shifts away from fuel-inefficient Cadillacs to fuel-efficient Toyotas, thereby raising the price of the latter relative to the price of the former such that more Cadillacs have to be given up in trade to obtain a Toyota. As before, general price levels are assumed constant. The resulting worsening of the U. S. terms of trade with Japan causes a real depreciation of the dollar relative to the yen necessitating an equilibrium shift in the exchange rate relative to the unchanged purchasing power parity.

Finally, suppose that Canada's export productivity doubles relative to U. S. export productivity, thereby rendering Canadian exports half as expensive as before in terms of U. S. exports. The resulting improvement in the U. S. real terms of trade causes the U. S. dollar to appreciate against the Canadian dollar despite no change in the purchasing power parity. The same thing would happen if productivity were advancing faster in the U. S. export sector than in the domestic nontradeables sector, thereby rendering exportables cheaper in terms of nontradeables. The resulting reduction in the real cost of transforming nontradeables into exportables and thereby into imports would strengthen the external value of the U. S. dollar on the foreign exchanges. In both cases the U. S. would experience a real exchange rate appreciation and a corresponding departure from purchasing power parity.

These examples illustrate how real shocks such as productivity growth differentials and international shifts in demand can induce systematic departures from purchasing power parity. In this perspective there is nothing mysterious about the failure of exchange rates to move in conformity with national price levels in the post-Bretton Woods era of floating exchange rates. Given the abundance of real shocks in that period (e.g., oil embargoes, commodity shortages, diverging productivity growth rates, shifts in international demands, changes in commercial policy and the like), persistent departures from purchasing power parity were to be expected.

Residual Validity of the Purchasing Power Parity Doctrine The last question raised by recent exchange rate experience refers to the residual validity of the purchasing power parity doctrine. What remains of the doctrine given the departures from purchasing power parity? Is it still a useful guide to exchange rate behavior? Does it still have something to teach us?

Regarding the validity of the doctrine the model presented above yields the following conclusions. First, the purchasing power parity doctrine can be expected to hold in the long run when the source of exchange rate disturbance is predominantly of a monetary origin. Such nominal disturbances have no lasting impact on the real exchange rate and therefore leave the purchasing power parity price-exchange rate relationship intact. Second, the doctrine is unlikely to hold in the short run since exchange rates tend to adjust to changes in underlying monetary conditions more quickly that national price levels thereby causing temporary divergences from purchasing power parity. Third, nor will purchasing power parity hold exactly in the long run when the source of exchange rate disturbance is of a predominately real origin. Such real shocks alter real relative prices and thereby loosen the linkage between price levels and exchange rates postulated by the doctrine. What this means is that although the purchasing power parity doctrine is a reliable guide to long run exchange rate movements originating in the monetary sector, its forecasting accuracy diminishes when real shocks affect exchange rates.

Note, however, that the purchasing power parity doctrine remains a useful tool even when real structural changes produce systematic disparities between exchange rate movements and changes in national price levels. It continues to be useful because it identifies divergent rates of price inflation as an important source of exchange rate movements and points out that this source could be eliminated if countries would pursue stable noninflationary monetary policies. That is, it specifies unstable monetary policies as a prime cause of exchange rate movements and stresses that this cause could be removed if countries would practice monetary stability. It also serves as a reminder that policies that strengthen the internal value of a currency will also strengthen its external value on the foreign exchanges. And, in stressing that exchange rates are endogenous variables determined by underlying monetary conditions, the doctrine yields the important insight that exchange rate depreciation per se cannot be inflationary because it merely reflects rather than creates underlying inflationary pressures. For these reasons the purchasing power parity doctrine remains a useful analytical tool.

Policy Implications Having employed the augmented monetary model to interpret recent exchange rate behavior, it remains to outline the policy implications of the model. At least three policy implications stem from the augmented monetary model. The first is that it is impossible for countries to peg both the external and internal value of their currencies in an inflationary world subject to real economic shocks. As a result, floating exchange rates may be necessary for countries desiring to stabilize general prices.

To see this, recall the real/nominal exchange rate equation $E = RP/P^*$. It is obvious from this expression that it is impossible to peg both the exchange rate E and domestic prices P if the real exchange rate R and/or foreign prices P^* are changing. Given foreign prices, a rising real exchange rate means that a country that wishes to stabilize its domestic price level must be prepared to abandon fixed rates and let its currency depreciate on the foreign exchanges. Similarly, given the real exchange rate, rising foreign prices mean that a country that wishes to pursue domestic price stability must let its currency appreciate on the foreign exchanges. Under these conditions only floating exchange rates are compatible with domestic price stability. For this reason, floating exchange rates may be necessary for countries wishing to achieve price stability in an inflationary world also subject to real economic shocks. Given price stability as the overriding policy goal, pegged exchange rates would be inferior to floating rates.

A second policy implication of the model is that proposed purchasing power parity policy intervention rules should be rejected. Such rules would require the authorities to intervene in the market for foreign exchange to insure that exchange rates conform to the purchasing power parity path of national price levels.

Such intervention rules are singularly ill-advised. They wrongly assume that policymakers can and should eliminate all real exchange rate changes. To be sure, real exchange rate changes stemming from erratic macroeconomic policies can and should be eliminated by pursuing stable, predictable policies. But real exchange rate changes stemming from fundamental structural changes in national economies should be accepted. For, as noted above, such changes generate changes in real relative prices that require equilibrium shifts in exchange rates relative to purchasing power parity. An intervention rule that ties exchange rates rigidly to national price levels in accordance with purchasing power parity ignores the need for such real exchange rate changes. It also fails to recognize that, because of sluggish national price levels, real exchange rate changes may be necessary to accomplish adjustments that would otherwise be achieved by movements in national price levels. For example, real exchange rate changes may serve the useful role of providing a temporary outlet for monetary shocks not accommodated by price level movements.

The model's third policy implication is that, short of obtaining a coordinated international program to equalize inflation rates in the trading world, the best way to reduce exchange rate fluctuations is to pursue a stable and predictable domestic noninflationary monetary policy. Not only would such a policy strengthen both the internal and external value of the currency, but it would also contribute to exchange rate stability in at least two ways: First, it would eliminate the unstable monetary growth that is a direct cause of exchange rate fluctuations. Second, it would exert a stabilizing effect on exchange rates via the expectations channel. This is so because market exchange rates are dominated by expectations of future monetary policies and these expectations themselves are influenced by current monetary policies. It follows that stable monetary policies will induce expectations of future policy stability and thereby exert a stabilizing influence on current exchange rates.

Note, however, that a stable domestic monetary policy alone would not eliminate all sources of exchange rate fluctuations. On the contrary, fluctuations could still result from unstable policies abroad as well as from unavoidable real disturbances. But stable monetary policy would eliminate one source of exchange rate variability, namely that produced by erratic and unstable domestic monetary policies. In short, while domestic monetary policy can do little to stop exchange rate disturbances originating from the real sector or from unstable policies abroad, it can stop one source of disturbance, namely that emanating from the domestic monetary sector.

Concluding Comments This article has presented an augmented monetary model of exchange rate determination and has used it to address certain questions raised by recent experience with floating exchange rates. The article's main conclusion is that real as well as monetary factors affect exchange rates and that they do so through nonmonetary channels, i.e., through real relative prices. The simplest version of the monetary approach ignores this, however, and for that reason must be augmented with a real exchange rate component if it is to account for all factors affecting exchange rates. Without this modification, the simple monetary approach is capable of accounting only for nominal movements in the exchange rate.

Even so, however, the simple version of the monetary approach remains a useful analytical tool. It provides a reliable guide to long-run exchange rate behavior when the source of exchange rate disturbance is of a predominantly monetary origin. It reminds us that excessive monetary growth is a primary source of exchange rate depreciation and that one can eliminate this source by adhering to stable noninflationary monetary policies. It also reminds us that exchange rates and price levels cannot be treated as independent, unrelated variables since policies that affect one tend to affect the other in the same way. It notes that since exchange rates themselves are determined by monetary policy they cannot be treated as independent policy instruments. And, in stressing that exchange rates are endogenous variables determined by underlying monetary conditions, it makes the important point that floating exchange rates cannot be inflationary since they reflect rather than generate inflationary pressures. In so doing, it effectively refutes the popular argument that floating rates cause inflation. These propositions remain valid even when real exchange rate changes occur. For this reason the monetary approach remains a useful analytical tool.

References

1. Artus, Jacques R., and Young, John H. "Fixed and Flexible Rates: A Renewal of the Debate." Working paper no. 367. NBER Working Paper Series. Cambridge: National Bureau of Economic Research, June 1979.

2. Bilson, John F. O. "The Current Experience with Floating Exchange Rates: An Appraisal of the Monetary Approach." *American Economic Review: Papers and Proceedings*, 68, no. 2 (May 1976): 392-7.

3. Flood, Robert P. "Explanations of Exchange-Rate Volatility and Other Empirical Regularities in Some Popular Models of the Foreign Exchange Market." Working paper no. 625. NBER Working Paper Series. Cambridge: National Bureau of Economic Research, February 1981.

4. Frenkel, Jacob A. "The Collapse of Purchasing Power Parities During the 1970s." Working paper no. 569. NBER Working Paper Series. Cambridge: National Bureau of Economic Research, October 1980.

5. ————. "Flexible Exchange Rates, Prices and the Role of 'News': Lessons from the 1970's." Report no. 8020. Center for Mathematical Studies in Business and Economics, University of Chicago, June 1980.

6. ————. "A Monetary Approach to the Exchange Rate: Doctrinal Aspects and Empirical Evidence." *Scandinavian Journal of Economics*, 78, no. 2 (May 1976): 200-24.

7. Kreinin, Mordechai E., and Officer, Lawrence H. *The Monetary Approach to the Balance of Payments: A Survey*. Princeton Studies in International Finance, no. 43. Princeton: Princeton University, International Finance Section, 1978.

8. Mussa, Michael. "The Exchange Rate, the Balance of Payments and Monetary and Fiscal Policy under a Regime of Controlled Floating." *Scandinavian Journal of Economics*, 78, no. 2 (May 1976): 229-48.

9. Sakakibara, Eisuke. "Purchasing Power Parity and Currency Substitution." *The Economic Studies Quarterly*, 30, no. 3 (December 1979): 202-18.

10. Stockman, Alan C. "A Theory of Exchange Rate Determination." *Journal of Political Economy*, 88, no. 4 (August 1980): 673-98.

59

SBCI

Swiss Bank Corporation Investment banking

JAPAN

Equity Research

WHARTON REPROGRAPHICS

March 16, 1988

Japan Economic Report

The Yen and Purchasing Power Parity

Summary

As of December 1987, the following six major currencies were all undervalued against the yen on a Purchasing Power Parity (PPP) basis:

US Dollar	-25%
Pound Sterling	-13%
Deutschemark	-29%
Swiss Franc	-2%
French Franc	-11%
Italian Lira	-8%

Examining the period from January 1976 to December 1987, we find that the maximum deviation (overvaluation or undervaluation) occurred in December 1979, when the Swiss franc was 30% overvalued against the yen. We were able to identify 19 periods of severe (+/-18%) overvaluation or undervaluation, but found that almost all were of short duration. Because overvaluation acts as a tax on exports and as a subsidy to imports, identifying the current state of an exchange rate provides useful information to microeconomic- and macroeconomic-analysts.

Adrian E. Tschoegl

Amsterdam
(020)264 100
Frankfurt
(069)714010
Hong Kong
(5)8446388
London
(01)600-0844
Melbourne
(03)654 6255
New York
(212)223-3007
San Francisco
(415)788 4996
Tokyo
(03)595-4300
Zurich
(01)223 11 11

Background

Purchasing Power Parity (PPP) is the theory that the exchange rate between two currencies will move to offset differences in relative inflation rates between the two countries. For example, if there is no inflation in Japan over a period of time, but high inflation in Italy over the same period, the theory predicts that the yen/Lit 100 exchange rate will move so that ¥1,000 will buy the same amount of Italian goods at the end of the period as it did at the beginning. (Please see the appendix for an algebraic statement of the theory and a stylized example.)

PPP is not a new theory. The theory of PPP is about 400 years old and tends to be rediscovered whenever inflation differentials between two major countries and currencies become extreme.

PPP has many limitations as a theory of exchange rates. The theory does not include the effect of changes in real factors such as technology, taste and resource supplies, nor in policy factors such as taxes, tariffs and non-trade barriers to trade. Also PPP is a long-term theory. Relative prices react sluggishly to changes in monetary conditions, but exchange rates react sharply; hence, over short periods of time, foreign exchange rates may deviate sharply from PPP values.

PPP charts and calculations can nevertheless be useful. PPP provides a broad guide to currency trends over the medium to long term in the case of large differences in the inflation rate histories of the countries involved. PPP provides a guide to which currencies are overvalued or undervalued in terms of inflation differentials.

Overvaluation or undervaluation have both microeconomic and macroeconomic implications. The reason is that an overvalued exchange rate acts as a tax on exports and a subsidy to imports. For example, if the PPP rate for the yen vis-a-vis the US dollar is ¥171/US$ 1, but the market rate is ¥128/US$ 1, the Japanese exporter gets ¥128 for a basket of goods which would sell in Japan for ¥171. The Japanese importer pays ¥128 for goods which he can sell at home for ¥171.

At the microeconomic level, knowing the approximate degree of over or undervaluation provides the analyst with a guide to one source of price pressure on import and export competing industries in the two countries. For example, in the case of the current yen/US dollar rate, are Japanese auto industry margins great enough that the companies can absorb the equivalent of a 25% sales tax without selling below variable cost, or will the companies have to raise prices in the US? The tendency for PPP to hold in the long run also provides the analyst with a way of estimating the probability of capital gains on a currency at the end of some period.

At the macroeconomic level, the tendency for PPP to hold in the long run means that deviations from fundamental trends are ultimately self-correcting. That is, trade balances and the exchange rate will respond to remove the deviation. However, the adjustments can take time both to begin and to run their course.

Because deviations from PPP represent overvaluations or undervaluations of a currency and, therefore, have real effects on the economies involved, governments have an incentive to attempt to influence the process. Central bank intervention affects the adjustment in two ways.

Intervention draws out adjustment over a longer time than freely functioning market forces would require. Central banks have chosen to take losses on their holdings of reserves rather than on government expenditures for welfare payments and stimulative programs to combat unemployment due to strong domestic currencies. Phased-in changes give firms time to adjust and result in less damage than do sudden changes. Central bank intervention also shifts adjustment to those currencies whose Central banks have been least willing to support their currencies.

READING THE CHARTS

Each of the six foreign exchange rate charts below graph the movement of the yen vis-a-vis another major currency. In each graph, the solid line depicts the actual exchange rate. To the right of each graph we give the maximum and minimum values which the actual exchange rate attained over the period the charts cover, i.e., January 1976 to December 1987. The parallel dotted lines in each chart give the PPP exchange rate, plus or minus 5%.

We have also designated points of extreme overvaluation and undervaluation, including the actual and PPP rates for that period and the degree of overvaluation or undervaluation. When the actual exchange rate is above the PPP band, we describe the foreign (i.e., non-yen) currency as being overvalued, and when the actual exchange rate is below the PPP band we describe the foreign currency as undervalued. Our formula for calculating the degree of overvaluation or undervaluation is the ratio of the actual exchange rate to the expected exchange rate given PPP, minus 1, and expressed in percent terms.

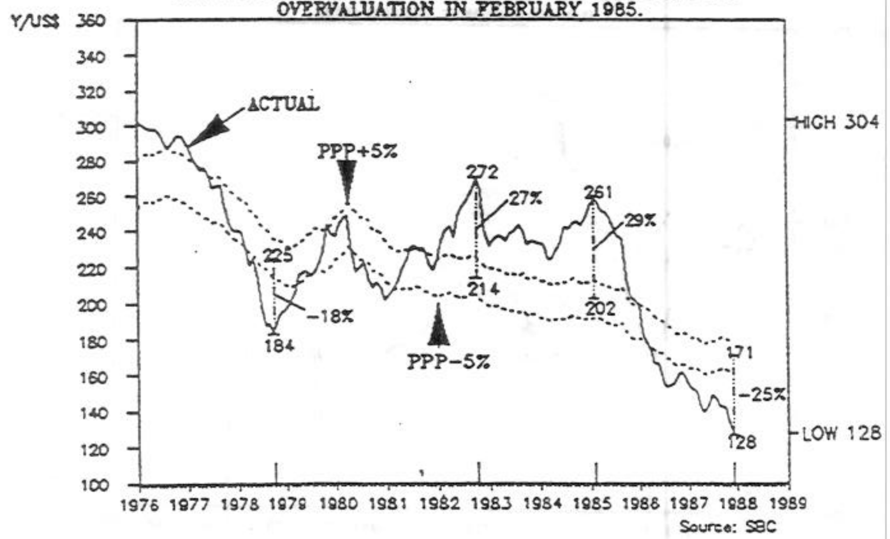
The data for the PPP exchange rates is the result of the most recent calculations from Dr. Erwin Heri of the Economics Division of Swiss Bank Corporation, Basle. The foreign exchange rate we call the actual rate is a monthly average of daily market rates. The PPP rates are the average of the results from two base periods, January 1973 to December 1973 and January 1979 to December 1984.

THE CURRENCIES

US Dollar

The US dollar reached its point of maximum deviation against the yen in February 1985 when it was 29% overvalued. More recently, the US dollar was 25% undervalued against the yen as of December 1987. Over the 12-year period, the US dollar has spent 23 months (16% of total time) in extreme (more than 18%) overvaluation or undervaluation. The US dollar remains undervalued when we change the inflation index and the starting date which we use for our calculations.

THE Y/US$ FOREX RATE HAS FOLLOWED THE PPP RATE DOWN.
THE POINT OF GREATEST DEVIATION WAS THE US$'S 29% OVERVALUATION IN FEBRUARY 1985.



Source: SBC

In our January 1988 Quarterly Economic Review, we argued that PPP calculations using wholesale prices and a 1978-1982 base period suggested that at ¥120/US$ 1, the US dollar was 30% undervalued. Therefore, the ¥120/US$ 1 rate represented something of a floor.

The common question with this or any other PPP calculation is, whether changing the index of relative inflation rates or the starting date which we use for our calculations would change the answer. From the table below, we can see that at ¥120/US$ 1, the dollar is substantially undervalued whether we measure from 1971, 1973, 1975 or 1980, and whether we use wholesale prices or the implicit GNP deflator.

The US Dollar Remains Substantially Undervalued at ¥120/US$ 1 Even When We Change The Base Date and the Price Index

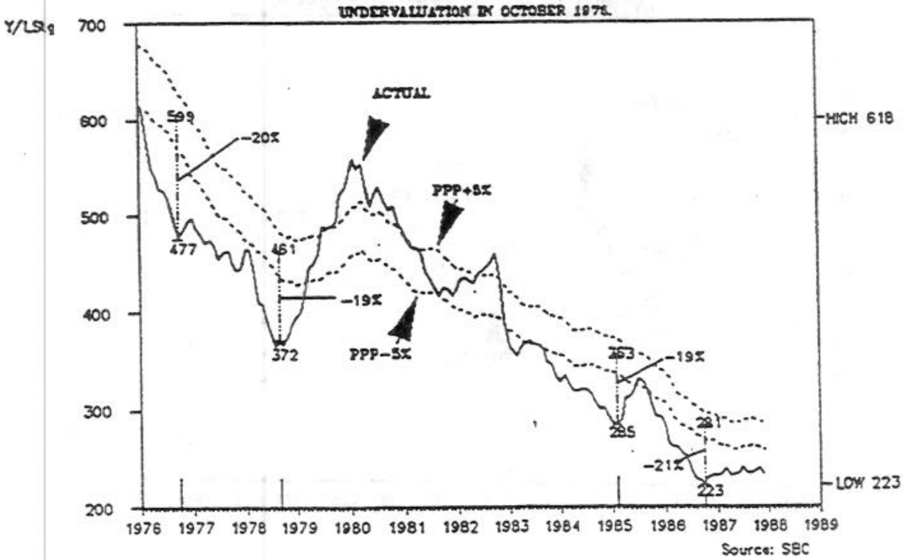
Base Date	WPI	GNP Deflator
1971	43%	53%
1973	37%	44%
1975	40%	44%
1980	23%	27%

Pound Sterling

The Pound Sterling was 13% undervalued against the yen in December 1987. The point of maximum deviation from PPP occurred in October 1986 when the Pound Sterling was 21% undervalued against the yen. The Pound Sterling spent 12 months (8% of total time) in extreme overvaluation or undervaluation.

There is one striking feature which is present in all the graphs to a greater or lesser degree. During the period from mid-1978 to late-1981, the PPP band zigzagged twice. For all six exchange rates, the actual exchange rate turned before the PPP rate turned. The 1979-1980 oil shock fed through to exchange rates immediately, and then to wholesale prices in general.

THE Y/LSTG FOREX RATE HAS FOLLOWED THE PPP DOWN. THE POINT OF GREATEST DEVIATION WAS LSTG'S 20% UNDERVALUATION IN OCTOBER 1976.

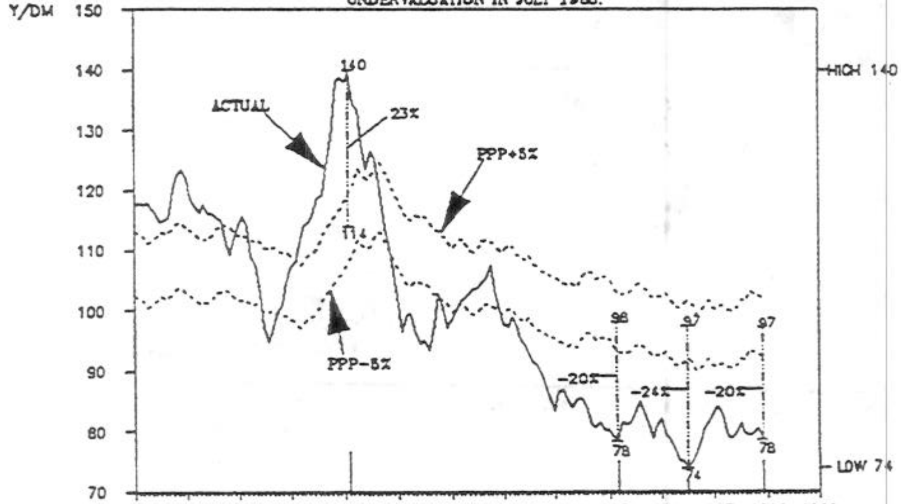


Source: SBC

Deutschemark

The Deutschemark was 20% undervalued against the yen in December 1987. The point of maximum deviation from PPP occurred in July 1986 when the Deutschemark was 24% undervalued against the yen. The yen/Deutschemark rate has been quite stable and, as a result, much of the movement of the exchange rate has been as a consequence of factors other than relative inflation differentials. The long period of 15-20% Deutschemark undervaluation from 1984 to the present is striking. Overall, the Deutschemark has spend 23 months (16% of total time) in extreme overvaluation or undervaluation.

THE Y/DM FOREX RATE HAS FOLLOWED THE PPP RATE DOWN.
THE POINT OF GREATEST DEVIATION WAS THE DM'S 24% UNDERVALUATION IN JULY 1986.

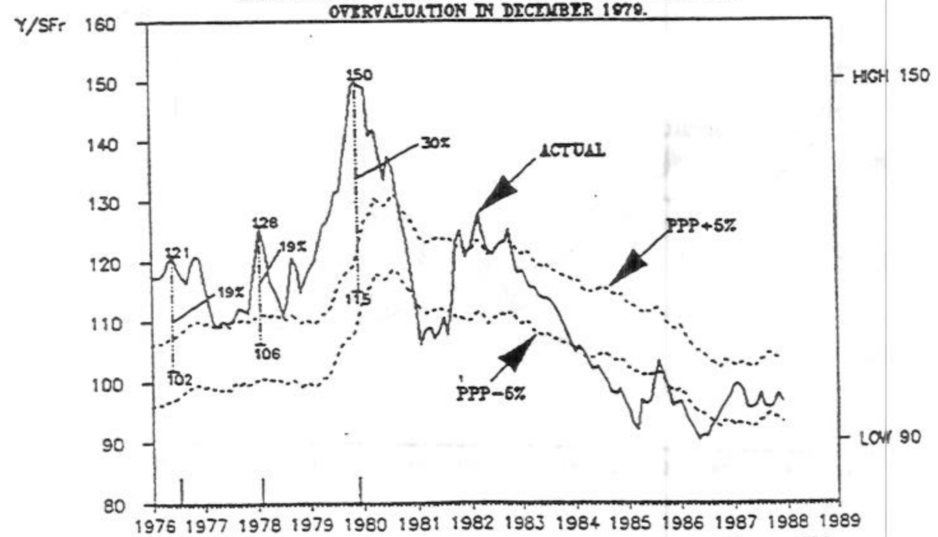


Source: SBC

Swiss Franc

The Swiss franc was 2% undervalued against the yen in December 1987. The point of maximum deviation from PPP occurred in December 1979, i.e., during the second oil shock, when the Swiss franc reached a level of 30% overvaluation against the yen. Since 1981, the yen/Swiss franc exchange rate has followed the PPP rate more closely than it did in the years before 1981. The Swiss franc has spent 13 months (9% of total time) in extreme overvaluation or undervaluation.

THE Y/SFR FOREX RATE HAS MOVED WITH THE PPP RATE.
THE POINT OF GREATEST DEVIATION WAS THE SFR'S 30% OVERVALUATION IN DECEMBER 1979.

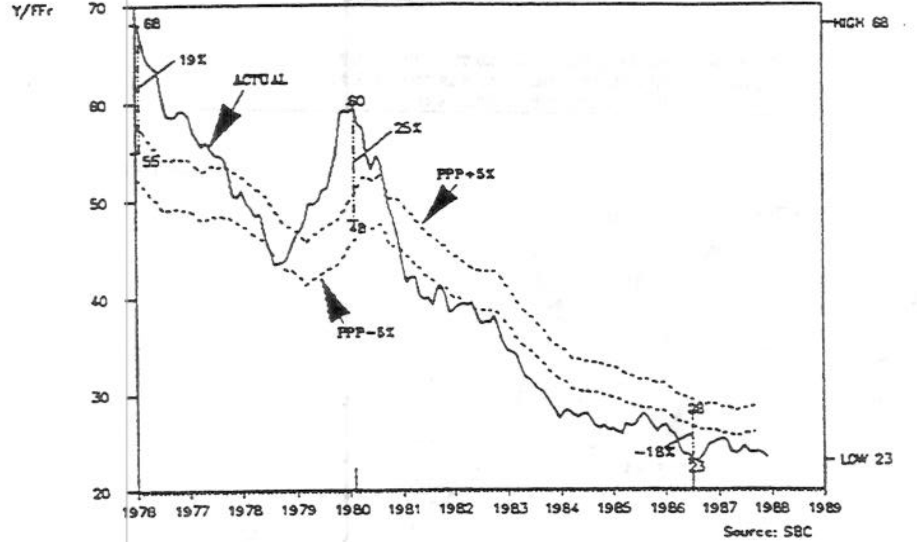


Source: SBC

French Franc

The French franc was 11% undervalued against the yen in December 1987. The point of maximum deviation from PPP occurred in February 1980 when the French franc was 25% overvalued against the yen. The French franc has been undervalued against the yen for almost the entire period from 1981, and has spent 13 months (9% of total time) in extreme overvaluation or undervaluation.

THE Y/FFR FOREX RATE HAS FOLLOWED THE PPP RATE DOWN.
THE POINT OF GREATEST DEVIATION WAS THE FFR'S 25% OVERVALUATION IN FEBRUARY 1980.

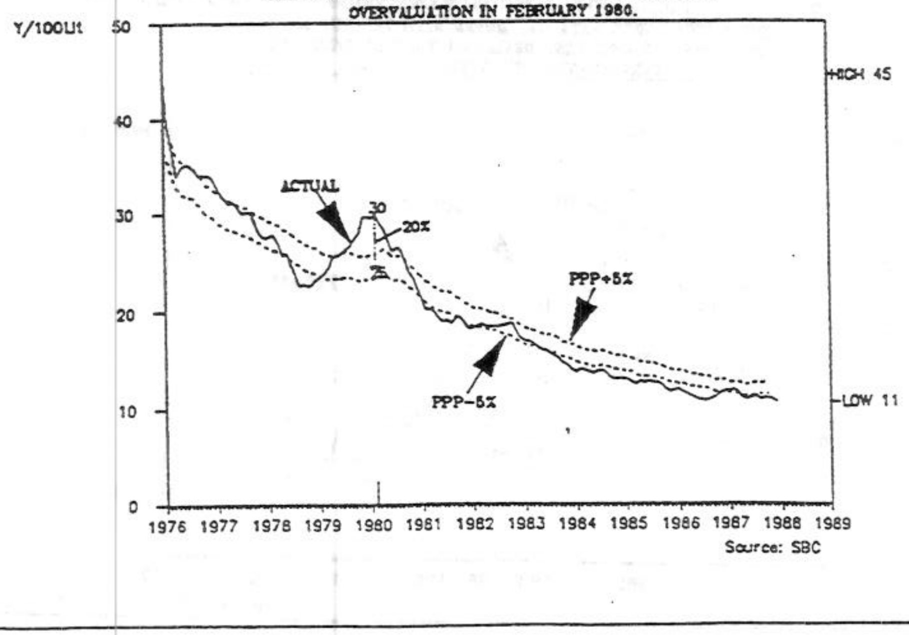


Source: SBC

Italian Lira

The Italian lira was 8% undervalued against the yen in December 1987. The point of maximum deviation from PPP occurred in February 1980 when the Italian lira was 20% overvalued against the yen. Except for the oil shock period, the deviations from the PPP band appear to have been minor. From a yen perspective then, Italian lira investments would appear to have had little foreign exchange risk. Almost all of the movement in the exchange rate has been due to relative inflation rates, which will have been reflected in interest or yield differentials. The relative competitiveness of the two countries' goods has not been subject to wild swings due to exchange rate changes. This also shows up in the fact that the lira spent only four months (3% of total time) in extreme overvaluation or undervaluation.

THE Y/LIT FOREX RATE HAS FOLLOWED THE PPP RATE DOWN.
THE POINT OF GREATEST DEVIATION WAS THE LIT'S 20% OVERVALUATION IN FEBRUARY 1980.



Source: SBC

SBCI

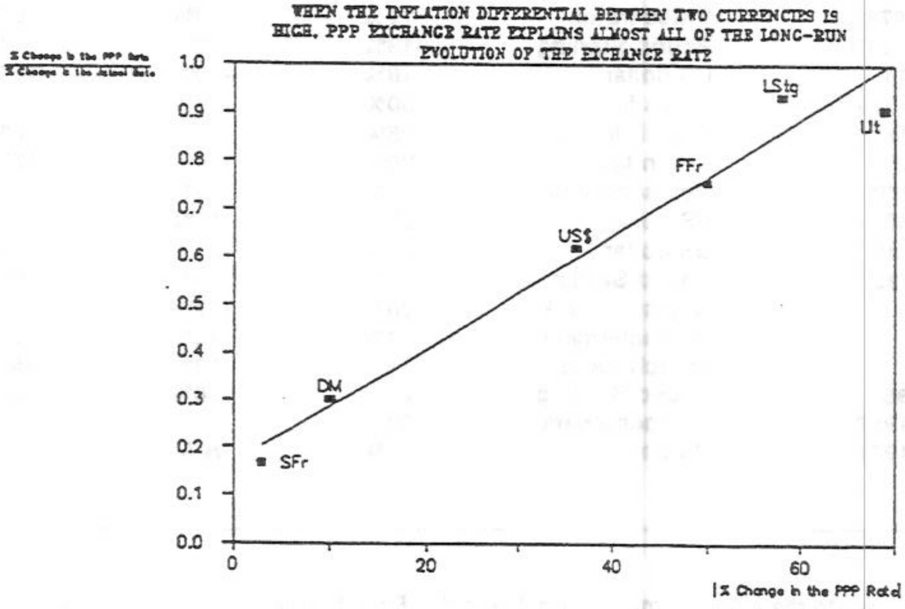
GENERALIZING FROM THE GRAPHS

In this section we will address two issues. The first issue is how well PPP (i.e., the inflation differential) describes the evolution of exchange rates. That is, when, if ever, is PPP a useful guide in the long run. The second issue is how persistent are deviations from PPP. This issue addresses the short-term utility of PPP notions.

PPP IN THE LONG RUN

Looking at the graphs, one can get a sense that PPP's utility as a theory varies quite dramatically from one currency to another. Clearly, there is a tendency for the exchange rate to evolve in the direction given by relative inflation rates. But the importance of the relative inflation factor varies. PPP seems to describe the evolution of the yen against the Italian lira, Pound Sterling and the French franc very well. PPP is moderately descriptive of the evolution of the yen/US dollar exchange rate. Finally, PPP is of only slight help in understanding how the Deutschemark and the Swiss franc have moved against the yen.

The graph below reinforces the intuitive impression. We have graphed the absolute value of the percentage change in the PPP exchange rate between January 1976 and December 1987 (the X-axis), against the ratio of that change to the percentage change in the actual foreign exchange rate over the same period (the Y-axis). What we have is a graph of how much of the twelve-year percentage change in the foreign exchange rate can be accounted for by relative inflation differentials, as a function of the magnitude of the inflation differential. The straight line in the graph is the regression line. (The correlation between our two variables is 0.99.)



WHEN THE INFLATION DIFFERENTIAL BETWEEN TWO CURRENCIES IS HIGH, PPP EXCHANGE RATE EXPLAINS ALMOST ALL OF THE LONG-RUN EVOLUTION OF THE EXCHANGE RATE

What the graph shows is that when the inflation differential between two currencies is high, PPP explains almost all of the long-run evolution of the exchange rate. However, when two currencies have very similar inflation histories, the relative contribution of the factors which PPP theory ignores becomes much greater.

PPP IN THE SHORT RUN

We can also see from the foreign exchange graphs that the actual exchange rate often deviates from the PPP rate, sometimes quite substantially. The largest deviation in our data was the Swiss franc's 30% overvaluation in December 1979 during the second oil shock.

One can think of an exchange rate, such as the yen/US dollar rate, as a drunk trying to follow a road in a valley. The PPP rate outlines the road, and the road's path is the result of the evolution of the US and Japanese inflation rates. The drunk can meander quite freely about either side of the road. However, the further away from the road he gets, the steeper the sides of the valley become, and the more likely that the drunk's next step will be back towards the road rather than away from it. Thus, overvaluation or under valuation of a currency on a PPP basis can persist for many years, but it is unlikely to remain extreme for long periods of time.

In the table below, we list the months in which any of our six exchange rates deviated from PPP by 18%, or more, for the first time in the subsequent six months. We then list the amount of the deviation and the amount of the deviation from PPP three months later.

MAJOR DEVIATIONS FROM PPP TEND NOT TO PERSIST

Date	Currency	Deviation (a)	Deviation 3 Months Later (b)	Change (|b| - |a|)
January 1976	French franc	24%	19%	-5%
June 1976	Swiss franc	19%	12%	-7%
October 1976	Pound Sterling	-20%	-7%	-13%
February 1978	Swiss franc	19%	9%	-10%
September 1978	Pound Sterling	-19%	-15%	-4%
October 1978	US dollar	-18%	-10%	-8%
December 1979	Swiss franc	30%	16%	-14%
February 1980	French franc	25%	10%	-15%
February 1980	Italian lira	20%	8%	-12%
February 1980	Deutschemark	23%	9%	-14%
October 1982	US dollar	27%	11%	-16%
February 1985	US dollar	29%	27%	-2%
February 1985	Pound Sterling	-19%	-9%	-10%
March 1985	Deutschemark	-20%	-17%	-3%
July 1986	Deutschemark	-24%	-18%	-6%
July 1986	French franc	-18%	-13%	-5%
October 1986	Pound Sterling	-21%	-16%	-5%
December 1987	Deutschemark	-20%	N.A.	
December 1987	US dollar	-25%	N.A.	
Average				-8.7%

Several points worth mentioning emerge from the table. First, for the seventeen extreme points for which we have three months of subsequent data, the deviation from PPP was always less extreme three months later. In fact, on average, the overvaluation or undervaluation had fallen by 8.7 percentage points. Extremes appear to be spikes, not plateaus. Our analogy of the drunk, the road and the sides of the valley appears to be a useful description of the behavior of exchange rates.

A second point worth mentioning is that there is no indication that our drunkard prefers one side of the road to the other. In all cases, the actual exchange rate has crossed the road several times. Moreover, we have nine cases of peak overvaluation and ten cases of peak undervaluation.

USING PPP FOR FORECASTING

This exchange rate history provides some guidance to the behavior the analyst can expect of exchange rates in the future. Given assumptions about the probability, in general, of extreme overvaluations or undervaluations and of the magnitude of expected inflation differentials, the analyst can readily construct forecasts of the possible future distribution of the exchange rate around its PPP value. These forecasts, in turn, permit inferences about exchange rate influences on economies and companies.

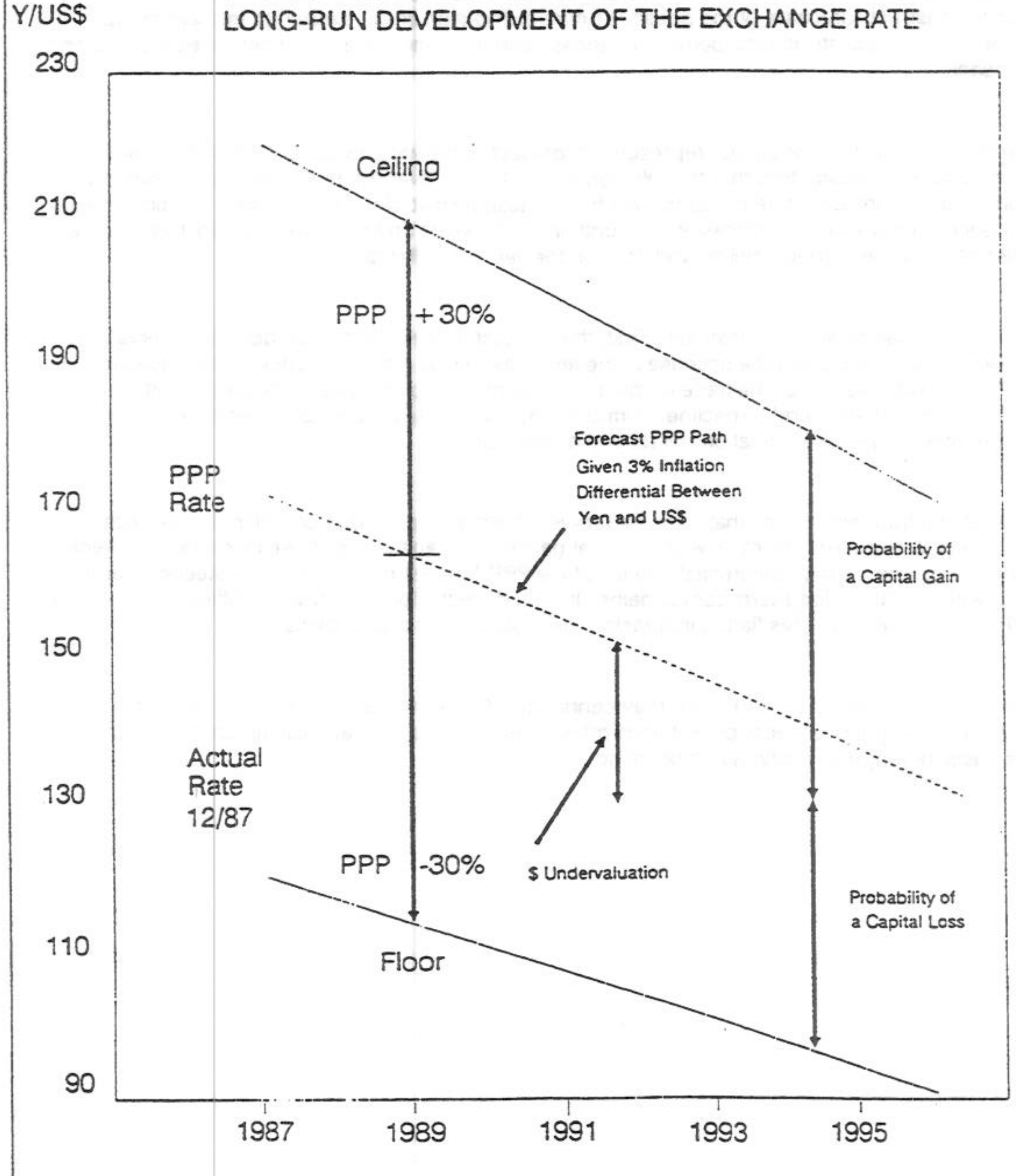
The graph on the following page represents a forecast of the yen/US dollar PPP rate ten years forward under the assumption that the inflation rate in the US over the period will be 3% per annum higher than the inflation rate in Japan. We further assume that 30% overvaluation and undervaluation is sufficiently rare (one occasion or month in our 12 years and six currencies) that we can treat those extremes as a rough ceiling and floor on the yen/US dollar rate.

The forecast has several implications. First, the forecast implies that the US dollar's undervaluation will tend to dissipate over time because of the actual exchange rate's tendency to move towards the PPP rate. This means that the trade competitiveness of US exports, due to the subsidy effect of undervaluation, will also tend to decline. Similarly, the price disadvantage of Japanese goods due to the tax effect of yen overvaluation will tend to decline too.

Second, the forecast implies that over the ten-year horizon, a purchase of dollars at the December actual rate is more likely to meet with a capital gain than a capital loss. If we increase our forecast of the size of the inflation differential, the fall of the PPP forecast rate will become steeper, decreasing the probability of long-term capital gains. If we reduce the forecast inflation differential, then the PPP forecast line becomes flatter, increasing the probability of capital gains.

The implications are not guarantees; they cannot be. All the implications are is an interpretation of the data in the light of a theory of exchange rates. Nevertheless, PPP and its implications do provide one basis for long-run analysis and planning.

THE ANALYST CAN USE PPP TO FORECAST THE LONG-RUN DEVELOPMENT OF THE EXCHANGE RATE



SBCI

Appendix on Purchasing Power Parity (PPP)

Definitions:

$(Y/\$)_o$ = Exchange Rate at some Base Date
$(Y/\$)_t$ = Exchange Rate Today
$(WPI)_{j,o}$ = Wholesale Price Index in Japan at the Base Date
$(WPI)_{j,t}$ = Wholesale Price Index in Japan Today
$(WPI)_{us,o}$ = Wholesale Price Index in US at the Base Date
$(WPI)_{us,t}$ = Wholesale Price Index in US Today

Theory:

$$(Y/\$)_t = (Y/\$)_o * [(WPI)_{j,t}/(WPI)_{j,o}] * [(WPI)_{us,o}/(WPI)_{us,t}]$$

Assumptions:

No inflation in Japan so:

$(WPI)_{j,o} = (WPI)_{j,t} = 100$

100% inflation in US so:

$(WPI)_{us,t} = 200$

and

$(WPI)_{us,o} = 100$

$(Y/\$)_o = 200$

Conclusion:

$(Y/\$)_t = 200 * (100/100) * (100/200) = 100$

Explanation:

At time t (e.g., today), the exchange rate in the example has moved to offset the inflation which occurred in the US since the base date (e.g., ten years ago) and which was not matched by a similar inflation in Japan. Because the exchange rate changed, ¥100 buys the same amount of US goods at the end of the period as it did at the beginning. Although ¥100 buys twice as many dollars at the end of the period as it did at the beginning, each dollar buys half as much today as it did ten years ago.

60

WHARTON REPROGRAPHICS

Big MacCurrencies

The dollar no longer looks cheap compared with the other big currencies

AS THE yen plunges and sterling soars, economists are being forced to revise their currency forecasts. To help them get their teeth into the subject, *The Economist* has updated its Big Mac index, which seeks to make exchange-rate theory a bit more digestible.

The Big Mac index is based upon the theory of purchasing-power parity (PPP), the notion that a dollar should buy the same amount in all countries. Supporters of PPP argue that in the long run, the exchange rate between two currencies should move towards the rate that would equalise the prices of an identical basket of goods and services in each country.

Our "basket" is a McDonald's Big Mac, produced in 110 countries. The Big Mac PPP is the exchange rate that would leave hamburgers costing the same in America as abroad. Comparing actual rates with PPPs signals whether a currency is under- or overvalued. (For more details on the index, check www.economist.com.)

The first column of the table shows local-currency prices of a Big Mac; the second converts the prices into dollars. The cheapest Big Macs are now in Indonesia and Malaysia, where they cost $1.16. At the other extreme, Big Mac fans in Switzerland have to pay $3.87. Given that Americans in four cities pay an average of $2.56, the rupiah and the ringgit look massively undervalued, the Swiss franc massively overvalued.

The third column calculates Big Mac PPPs. For example, dividing the Japanese price by the American price gives a dollar PPP of ¥109. On April 6th, the exchange rate was ¥135, implying that the yen is 19% undervalued against the dollar. Three years ago the index suggested that the yen was 100% overvalued against the dollar. Likewise, the D-mark is now only 5% overvalued, against 50% in April 1995.

Thanks to the dollar's rise—long predicted by burgernomics—it is now closer to its PPP against other big currencies than for many years. Indeed, only five currencies in the table are now significantly overvalued against the greenback, among them Britain's, Sweden's and Denmark's. All three countries have decided not to adopt Europe's single currency, the euro, next year. The pound is 19% overvalued against the dollar, which implies it is 14% overvalued against the D-mark. In contrast, the currencies of the euro-block countries are close to Mcparity against the D-mark.

The most dramatic changes in the index over the past year are in East Asia, where devaluations have left currencies significantly undervalued. This competitive advantage, however, is being eroded by inflation. In Indonesia, the price of a Big Mac has more than doubled over the past year. East European currencies also look cheap, with the Hungarian forint 52% undervalued against the dollar.

The Big Mac index is not a perfect measure of PPP. Price differences may be distorted by trade barriers on beef, sales taxes, local competition and changes in the cost of non-traded inputs such as rents. But despite its flaws, the Big Mac index produces PPP estimates close to those derived by more sophisticated methods. A currency can deviate from PPP for long periods, but several studies have found that the Big Mac PPP is a useful predictor of future movements—enabling the hungry investor to get rich by putting his money where his mouth is.

The hamburger standard

	Big Mac prices		Implied PPP* of the dollar	Actual $ exchange rate 6/4/98	Under (-)/over(+) valuation against dollar, %
	in local currency	in dollars			
United States†	**$2.56**	**2.56**	–	–	–
Argentina	Peso2.50	2.50	0.98	1.00	-2
Australia	A$2.65	1.75	1.04	1.51	-32
Austria	Sch34.0	2.62	13.28	12.96	+2
Belgium	BFr109	2.87	42.58	38.00	+12
Brazil	Real3.10	2.72	1.21	1.14	+6
Britain	£1.84	3.05	1.39‡	1.66‡	+19
Canada	C$2.79	1.97	1.09	1.42	-23
Chile	Peso1,250	2.75	488	455	+7
China	Yuan9.90	1.20	3.87	8.28	-53
Czech Republic	CKr54.0	1.57	21.1	34.4	-39
Denmark	DKr23.8	3.39	9.28	7.02	+32
France	FFr17.5	2.84	6.84	6.17	+11
Germany	DM4.95	2.69	1.93	1.84	+5
Hong Kong	HK$10.2	1.32	3.98	7.75	-49
Hungary	Forint259	1.22	101	213	-52
Indonesia	Rupiah9,900	1.16	3,867	8,500	-55
Israel	Shekel12.50	3.38	4.88	3.70	+32
Italy	Lire4,500	2.47	1,758	1,818	-3
Japan	¥280	2.08	109	135	-19
Malaysia	M$4.30	1.16	1.68	3.72	-55
Mexico	Peso17.9	2.10	6.99	8.54	-18
Netherlands	Fl5.45	2.63	2.13	2.07	+3
New Zealand	NZ$3.45	1.90	1.35	1.82	-26
Poland	Zloty5.30	1.53	2.07	3.46	-40
Russia	Rouble12,000	2.00	4,688	5,999	-22
Singapore	S$3.00	1.85	1.17	1.62	-28
South Africa	Rand8.00	1.59	3.13	5.04	-38
South Korea	Won2,600	1.76	1,016	1,474	-31
Spain	Pta375	2.40	146	156	-6
Sweden	SKr24.0	3.00	9.38	8.00	+17
Switzerland	SFr5.90	3.87	2.30	1.52	+51
Taiwan	NT$68.0	2.06	26.6	33.0	-20
Thailand	Baht52.0	1.30	20.3	40.0	-49

*Purchasing-power parity: local price divided by price in United States
†Average of New York, Chicago, San Francisco and Atlanta ‡Dollars per pound
Source: McDonald's

TOBIN, J. (1970). Money and income: *post hoc ergo propter hoc. Quarterly Journal of Economics*, 84, pp. 301–29.

TOOKE, T. (1844). *An Inquiry into the Currency Principle.* London: Longman, Brown, Green and Longmans.

WICKSELL, K. (1898). *Geldzins and Guterpreise (Interest Rates and Prices).* Jena: Fischer.

LIPPINCOTT LIBRARY
DOCUMENT DELIVERY

Scottish Journal of Political Economy, Vol. 43, No. 4, September 1996
© Scottish Economic Society 1996. Published by Blackwell Publishers Ltd., 108 Cowley Road, Oxford OX4 1JF, UK and 238 Main Street, Cambridge, MA 02142, USA

US FOREIGN EXCHANGE MARKET INTERVENTION SINCE 1962

*Anna J. Schwartz**

ABSTRACT

The Federal Reserve has intervened in foreign exchange markets since 1962 as a partner of the Treasury's Exchange Stabilization Fund. Departing from its scripted passive role under the Bretton Woods system, the US defense of the dollar under pegged exchange rates may not appear unreasonable. Why the practice continues under a floating rate system is harder to understand. By accumulating large foreign currency balances since 1987, the US has made intervention easier. Holding larger balances does not help the authorities know what exchange rates ought to be. The US for a few years in the 1980s refrained from intervening. It can happen again if its authorities acknowledge that intervening is an exercise in futility.

I INTRODUCTION

US attempts to stabilize exchange rates by relying on intervention in foreign exchange markets began under the Bretton Woods par value system. US intervention has continued since the move to generalized floating in the early 1970s despite conflicting views on its effectiveness. Opponents and advocates of intervention are found both among monetary authorities and academic economists. In this paper I examine the US position on intervention since 1962 when it took hold as an important policy.

In the United States two agencies conduct foreign exchange operations. The senior agency is the Exchange Stabilization Fund (ESF), a part of the Treasury Department that was established by the Gold Reserve Act of 1934 with authority to operate in the foreign exchange market, specifically, to protect the exchange value of the dollar. The Federal Reserve Bank of New York since 1961 has been the instrumentality that executes transactions for the Treasury's account at its request. Since 1962 the Federal Reserve Bank of New York has also been undertaking foreign exchange operations for the System's own account. The System is the junior agency and the decision that it should intervene is at the Treasury's initiative. In practice, the Federal Reserve and the Treasury coordinate their operations.

Although the Federal Open Market Committee assumed the authority to engage in exchange market operations without an express legislative mandate, Congress was not left in the dark about the Federal Reserve action in this matter. Congress,

WHARTON REPROGRAPHICS

837

however, neither in 1962 nor in any year since chose either to oppose or to ratify the Federal Reserve's action. Yet it did so obliquely in the Monetary Control Act of 1980 in Section 14(b) (1), which authorized System purchases of foreign government and government agency obligations for the purpose of investing idle funds—obviously the proceeds of foreign exchange market interventions.

The Treasury enlisted the Federal Reserve as its partner in foreign exchange market intervention under the Bretton Woods system. Beginning in 1961 the Exchange Stabilization Fund intervened to stem the drain of the US monetary gold stock as countries converted dollars they had accumulated in their official reserves. With limited resources to engage in intervention—ESF capital had been transferred in 1947 in partial payment of the US quota in the International Monetary Fund (IMF)—the Treasury persuaded the Federal Reserve to supplement its intervention efforts.

One of the anomalies of the decision by the Federal Reserve to become the Treasury's partner is that no record exists before 1990 of a discussion by the Federal Open Market Committee (FOMC), the System's chief policy-making arm, of the relation of exchange market intervention to monetary policy. The FOMC debated at several meetings in 1961–62 whether the System had legal authority to intervene but, so far as the record shows, it did not debate the domestic monetary policy implications of intervention.[1]

A subcommittee of the FOMC, headed by the chairman, after consultations with the Treasury, decides the amount of the System's purchases or sales of individual currencies and instructs the manager of the System foreign account at the New York Federal Reserve Bank to execute the transaction. The full FOMC does not discuss intervention in advance. The manager merely asks the FOMC for a vote ratifying his intervention. No vote is cast against the manager's report since he is only carrying out the chairman's instructions. If there is any discussion, it does not appear in the record. Dealings with foreign central banks are regarded as privileged.

In what follows I assess the record of US intervention to date. Section II examines the record under Bretton Woods and since the transition to the floating exchange rate regime. Section III discusses the relationship between sterilized intervention and US exchange rates. Section IV tries to determine the rationale for such intervention when the evidence suggests that it has little effect on exchange rates and monetary policy. Section V concludes.

II RECORD OF INTERVENTION

Bretton Woods era

The Bretton Woods Agreement in 1945 envisioned that short-run payments imbalances would be met by drawings on official exchange reserves and IMF

---

[1] At the FOMC 27 March, 1990, meeting, the participants had a full-dress discussion of intervention in foreign exchange markets and of the legality of warehousing. The main subject was not the monetary implications of intervention operations, although there were incidental observations about sterilized vs. unsterilized intervention. The main concern of the participants was the growing size of foreign currency balances held by the Federal Reserve.

© Scottish Economic Society 1996

credits to ensure that the economy would be insulated from external shocks. This would imply that the domestic monetary impact of exchange-market interventions would be sterilized, and that the US along with other countries would have the option to change its exchange rate (Bordo, 1993; McKinnon, 1993). In operation, however, the Bretton Woods system had one set of rules for countries other than the US and another set for the US The US became the centre country, not obligated to intervene directly in foreign exchange markets, as were other industrialized countries, which found that only unsterilized intervention that changed their domestic monetary bases was effective. Since they invested official reserves in interest-earning US Treasury securities these could have grown without threat to the US gold stock, provided that they did not attempt to convert their holdings into gold. The mostly voluntary growth of dollar claims relative to the size of the gold stock, however, led to the collapse of the dollar-based par value system, given the pro forma commitment by the US to convert dollars into gold at $35 an ounce and US retreat from conservative monetary policy after the mid-1960s. Without gold convertibility, the system could have continued indefinitely as a dollar standard. A detailed examination follows of the steps taken by the US because of dissatisfaction with its situation under the Bretton Woods system.

Intervention in 1962, according to the FOMC, was designed to (a) offset pressure on the US monetary gold stock when deemed to be temporary (reversible); (b) smooth (offset) disequilibrating movements in exchange rates, i.e., stave off speculative attacks and prevent disorderly exchange markets; (c) supplement international arrangements through the IMF; and (d) satisfy the world demand for liquidity. Item (d) suggests that intervention was not intended to affect the deficit in the US balance of payments; and that the US would continue to supply dollars in payment for the goods and services which the rest of the world provided.

Intervention forms were varied, all aimed at giving the US foreign currency balances:

(i) The basic form was swap lines—lines of credit established with the central banks of the leading European countries, to provide for the right of the US to draw on a foreign currency up to a specified ceiling for a specified term, and the reciprocal right of the foreign central bank to draw on US dollars at the current exchange rate for a matching sum and term. Except for the Bank of Mexico, the other central banks with which the Federal Reserve had swaps were in major industrialized countries. Treasury reciprocal currency agreements during the Bretton Woods period were with central banks of less advanced countries.

(ii) The second form was a US drawing on its IMF quota, by which it purchased foreign currencies for dollars, payable over a three to five-year period.

(iii) The Treasury persuaded the IMF and the other members of the G-10 in December 1961 to create a supplemental facility, the General Arrangements to Borrow (GAB). It provided the Fund with a $6 billion line of credit from central banks in balance-of-payments surplus to assist countries in deficit, in particular, the United States. The US quota in the IMF was nearly $6 billion whereas the

© Scottish Economic Society 1996

IMF held only about $1.5 billion in currencies other than dollars. Hence the United States could not draw enough in foreign currencies from the IMF to meet its reserve needs. The GAB was intended to serve as a supplementary source of liquidity, but it did not however, correct the illiquidity problem of the IMF (Mundell, 1993, p. 606).

(iv) The Treasury sold securities denominated in foreign currencies which are referred to as Roosa bonds, after the then Under Secretary of the Treasury for Monetary Affairs.

(v) The US encouraged prepayment of official foreign indebtedness—counterpart funds from Marshall Plan assistance to European countries.

(vi) The Federal Reserve warehoused foreign currencies which the Treasury lacked funds to acquire in its own. The origins of this practice may be traced to the Federal Reserve's purchase of $33.5 million from the ESF in 1962 for its opening balances in D-marks, Swiss francs, guilders, and Italian lira. The idea was to provide the Federal Reserve with the possibility of repaying foreign currencies initially acquired under swap lines, since the term of the facilities was then three to six-months. The original 1962 guidelines of the FOMC for System transactions with the ESF could be read as sanctioning warehousing, although the term was not then in use. The term was first used in 1963 by the then manager of the Trading Desk at the Federal Reserve Bank of New York, and warehousing occurred sporadically in the Bretton Woods years, but did not become significant until 1977 (Todd, 1992, p. 133).

(vii) The Federal Reserve and the Treasury intervened directly to buy and sell foreign currencies.

During the Bretton Woods era intervention was directed to protect the value of dollars held by foreign central banks in order to deter their exchange for US gold.

The principal form of intervention, as noted, was the arrangement with foreign central banks, called swap agreements, which allowed the Fed to obtain foreign currency equivalent to a specified dollar ceiling for a designated period. The funds could be rolled over, but were usually repaid in one or two quarters. If the Fed sold marks (bought dollars) one day, that would cause the value of the dollar relative to the mark to fall in the forward market, since traders knew that the marks would have to be repurchased in the future, but not when it would occur. Uncertainty about the future spot rate would increase as would uncertainty about the forward rate that day.

The United States borrowed foreign currencies to buy dollars held abroad rather than selling gold. To repay the swaps, the Treasury at the same time issued long-term (Roosa) bonds denominated in foreign currencies, thus avoiding a reduction in gold reserves. The Fed, however, could suffer losses under the swap arrangements. Its sales of a foreign currency could encourage traders to buy in anticipation of a rise in the price of the foreign currency when the Fed repurchased it to repay its borrowing.

As the decade progressed, the authorities intervened massively in every major currency, reflected in the growth of total Federal swap lines from $900 million equivalent in 1962 to $11,730 equivalent in 1971.

© Scottish Economic Society 1996

The first swap partner to draw on the reciprocal currency arrangement with the System was the Bank of Canada in June 1962, in defence of the Canadian dollar, Canada having returned to a fixed parity in May 1962. The Bank of England availed itself of the swap line in most quarters from 1963 on.

The defence of the dollar during the Bretton Woods period was a losing game. The reason was that the United States did not use the period of intervention to adopt policies to correct its balance of payments and the overvaluation of the dollar. It was inevitable that the gold-dollar system would collapse without a devaluation of the US monetary gold stock, even if the US had observed a stable monetary policy. A devaluation in that event would have been time-inconsistent (Meltzer, 1993). When the United States closed the gold window on August 15, 1971, it abandoned the game in order to achieve a dollar devaluation that would improve its trade balance.

Transition to floating

From August 1971 to March 1973, the authorities tried to preserve the fixed exchange rate system, and to defend the new parities established in December 1971 under the Smithsonian Agreement, when the official dollar price of gold was raised. The central rates crumbled during the nine months following the floating of sterling in June 1972. On February 10, 1973, Japan closed its foreign exchange market and suspended support of the dollar. New central rates were set in a hurried round of negotiations although many currencies floated. Again the official dollar price of gold was raised, but the new central rates did not staunch the flow of dollars abroad. When a further crisis erupted in March 1973, the major industrialized countries discontinued pegging their exchange rates to the dollar. Nevertheless, the dollar's role as the international reserve currency remained unchanged.

The US authorities continued the intervention policies of the 1960s. It is not clear, however, why they did so, since they were no longer subject to the constraints of the Bretton Woods system. In July 1973 the Federal Reserve intervened with its own small holdings of foreign currency, and with the much larger total of foreign currency resources available through swap agreements. The authorities were then concerned by the fall in the dollar as US inflation rose, the prospect of higher oil imports, and the political fallout from the Watergate incident. At the same time European currencies were strengthening, leading the US to intervene with modest sales of D-marks.

From the advent of generalized floating, until September 1974, the authorities intervened sporadically to counter what they regarded as disorderly markets. In May 1974, after several months of dollar depreciation, concerted intervention was undertaken by the US, German, and Swiss authorities. The exchange risk on US drawings on swaps was shared evenly with the foreign central banks on which drawings were made. The interest rates paid on any drawing either by the System, the Treasury, or by foreign central banks were based on current market rates for US Treasury bills.

© Scottish Economic Society 1996

Renewed Intervention, 1974–1981

Three episodes of intense intervention characterized the period:

(i) From October 1974 to March 1975, the Federal Reserve, the Bundesbank, and the Swiss National Bank coordinated intervention to offset volatility in the dollar-D-mark and dollar-Swiss franc rates and to slow depreciation of the dollar. The System's intervention, amounting to $1.4 million dollar equivalent, was financed by drawings on swap lines, the other central banks making somewhat larger dollar purchases.

(ii) From September 1977 to December 1979, the Bank of Japan joined the coordinated intervention team to arrest the decline in the dollar. The ESF reached a swap agreement with the Bundesbank on 4 January, 1978, to supplement the Bundesbank swap line with the System, which was doubled. In addition, the Treasury sold SDRs to the Bundesbank for marks. The Treasury auctioned gold to the public, drew on the US reserve position in the IMF, and issued bonds enominated in foreign currencies.

The decline in the weighted average exchange value of the dollar nevertheless accelerated in 1978. An anti-inflation program announced on October 24 (involving fiscal restraints, voluntary wage and price standards, and a reduction in regulatory actions) did not moderate the dollar's slide on the exchange market. On 1 November, foreign exchange resources equivalent to $30 billion were mobilized to finance intervention to support the dollar, in cooperation with Germany, Japan, and Switzerland. During the last two months of 1978, US exchange market intervention in support of the dollar totalled $6.7 billion and the three cooperating countries made significant purchases of dollars.

By June 1979 the dollar's value on a trade-weighted basis had risen by 10% from its 1978 low, and the US authorities had repurchased a greater sum of foreign currencies than they had sold in the last two months of 1978. The dollar then began to weaken, and US sales of foreign currencies, chiefly marks, resumed (Greene, 1983). On 6 October, 1979, the Federal Reserve announced a wide-ranging set of measures to tighten monetary control, and the dollar began to appreciate. From October 1979 on, the authorities intervened frequently, operating on both sides of the market. When the dollar was in demand, they acquired foreign currencies in the market and from correspondents to repay earlier debt and to build up balances. They were buyers from February to March 1980, and from late March to early April and beyond, they sold currencies. By the end of July, the authorities were again accumulating currencies.

The warehousing facility for the Treasury was officially approved by the FOMC in January 1977. It authorized purchases by the System of up to $1.5 billion of developed country currencies from the ESF for 12 months. In 1978 the FOMC authorized use of the warehousing facility for the Treasury's General Fund. The System warehoused proceeds of Carter bonds—bonds denominated in marks and Swiss francs issued in 1978–80—for the General Fund, that is, the System made a loan to the Treasury for the loans the Treasury had obtained abroad. The System was prepared to warehouse currencies with no time limits, subject only to the Treasury's annual settlement for valuation gains and losses.

© Scottish Eco Society 1996

(iii) By the end of 1980, the Federal Reserve was intervening in the foreign exchange markets virtually on a day-to-day basis. To offset disorderly markets, the Trading Desk placed simultaneous bid-and-asked prices. From September 1980 to February 1981, the aim of intervention was to cushion the rise in the dollar, and to acquire hard currencies to pay off swap debts. The authorities bought D-marks, Swiss francs, French francs, more than doubling their foreign currency assets.

Changes introduced at the beginning of 1981 included the agreement by the System, the Treasury, and their swap partners to take the full exchange risk on their drawings, and to pay a rate of interest based on the creditor country's Treasury bill rate or an equivalent market rate. These changes had no immediate effects since on 19 February, 1981, the Reagan Administration halted exchange market operations, intervening later that month to calm a disorderly market and thereafter only once at the time of the assassination attempt on President Reagan.

Little intervention, 1981–1986

During this period, intervention operations by the US authorities were limited. The support for market determination of foreign exchange rates in the first Reagan Administration was opposed by some European monetary authorities. At the Versailles Summit of the G–7 in June 1982, it was agreed to sponsor studies by a working group of the effectiveness of intervention in foreign exchange markets. These studies formed the basis for the Jurgensen report the following year, which concluded that intervention could influence exchange rates in the short run, but that sterilized intervention which left the monetary base unchanged had no lasting effect unless accompanied by domestic policy changes. The report did not alter the beliefs of European authorities with respect to the value of intervention. They continued to use the System's Trading Desk as their agent when they chose to intervene in the New York market.

The centerpiece of this period was the persistent rise in the market value of the dollar from 1981 until February 1985, and the accompanying debate on whether there was a bubble in the 1984–85 market valuation of the dollar, in view of the easing of monetary policy in mid–1984. For many economists and corporate officials from the real sector, this was a clear case of misalignment which was responsible for a widening of the US merchandise trade deficit and a heavy loss of jobs in the traded goods sector. Industrial countries resented the need to tighten domestic monetary policy as a counterweight to the upward pressure on the exchange value of the dollar.

In the second Reagan Administration, the new Secretary of the Treasury, James A. Baker III, retreated from the position favouring market-determined exchange rates. Although the dollar had begun to weaken, in February 1985, the US authorities, to reverse appreciation of the dollar, convened a meeting of G-5 officials at the Plaza Hotel in New York on September 22. In their view, 'exchange rates should play a role in adjusting external imbalances.' Following the Plaza Accord, massive intervention occurred by the G-5. They sold dollars (bought other currencies).

© Scottish Economic Society 1996

How much of the dollar's decline between February 1985 and December 1985 would have occurred in the absence of intervention is an open question. Edwin M. Truman, staff director of the Division of International Finance of the Federal Reserve System, concludes with respect to the January-March 1985 episode that 'intervention per se could not have had much to do with a turn in the dollar that was long overdue' (Kenen *et al.*, 1994, p. 250). He regards the September-November 1985 episode as 'partially successful'.

The US authorities did not intervene in 1986, since they regarded the continued decline of the dollar as orderly. The Japanese, however, bought dollar in the second and third quarters of the year to counter appreciation of the yen.

Intervention was apparently sterilized by the G-5. Coordination was evident between the Bundesbank and the Federal Reserve, but less so between them and the Bank of Japan (Weber, 1994). Intervention was ineffective measured by no reversal of trends in bilateral exchange rates.

The ESF during this period made swap loans to Argentina, Brazil, Mexico, the Philippines, Ecuador, Nigeria, and Yugoslavia. These were temporary currency stabilization loans for developing countries experiencing payments crises. Clearly, their problems did not impact the exchange value of the dollar or foreign exchange markets in general.

Renewed intervention, 1987–1994

During this period, marked by coordinated central bank intervention under G-7 prodding, US purchases of foreign currencies exceeded those of preceding years. Maximum amounts of authorized Federal Reserve foreign currency balances rose from $12 billion in 1987 to $25 billion in 1990, and actual balances rose accordingly. The Treasury's balances also increased *pari passu.*

In daily data between January 1987 and March 1991, researchers have identified 17 individual episodes of coordinated intervention , which they have interpreted as either temporarily successful—the next episode is of the same sign—or definitely successful—the next episode is of the opposite sign (Catte, Galli and Rebecchini, 1994). Doubts concerning the validity of these findings have been expressed (Truman, 1994; Weber, 1994).

At the end of January 1987, in coordination with Japan, the US authorities sold yen to counter depreciation of the dollar. The G-7 met at the Louvre in Paris on 22 February to foster stability of exchange rates at current levels. Yet the dollar continued to decline against the yen, despite intervention. The Federal Reserve's injection of liquidity into the economy in response to the stock market crash in October was accompanied by a further decline in the dollar, the Louvre Accord notwithstanding. Throughout 1987 the support of the dollar by major central banks amounted to nearly ten times the size of US intervention. In 1988 US net intervention was moderate as the authorities both bought and sold foreign currencies.

A stronger dollar in 1989 led to three weeks of coordinated intervention in October to lower its exchange value. US purchases of yen and marks in this episode were on a larger scale than in previous years, and sales of dollars by the G-7 were twice the size of US intervention. Warehousing by the System enabled the Treasury to acquire additions to its holdings.

In 1990 US authorities bought yen and marks. In the middle of the year, the ESF sold its mark balances in order to retire amounts in the warehouse, and the System sold towards the end of the year to counter a bolstered mark. These operations continued early on in 1991 with purchases of yen and marks, followed by spot and forward sales of marks. The pattern in 1992 was much like that in 1991. In 1993 intervention action was limited to the yen, with alternating sales and purchases. In the first half of 1994, the authorities sold both yen and marks. At the end of 1994 and during the first months of 1995, the fall in the foreign exchange value of the dollar against the yen and the mark sometimes led to an organized effort to support the dollar, at other times to passive acceptance by the US and other central banks of the market's verdict.

The swap line for Mexico was activated in 1988, 1990, 1992, 1994 and most notably in 1995, a loan of unprecedented magnitude and term was extended from the ESF to Mexico. These were actions to support the peso that had less to do with US economic than with its foreign policy interests.

III STERILIZED INTERVENTION AND EXCHANGE RATES[2]

Sterilized vs. unsterilized intervention

Official intervention by the Federal Reserve is sterilized intervention. When it purchases a given dollar equivalent of foreign currencies, it sells a corresponding dollar amount of US government securities. When it sells a given dollar equivalent of foreign currencies, it purchases a corresponding dollar amount of US government securities. Foreign exchange market intervention is sterilized, so that the monetary base that the Federal Reserve controls is unchanged by intervention. It neither seeks nor desires to expand or contract the base as a result of intervention.

It is generally accepted that unsterilized intervention by central banks does affect exchange rates. Unsterilized intervention is monetary policy conducted by open market operations in foreign securities rather than domestic securities. A change in the relative size of the monetary base, however achieved, compared to the change in the base in other countries can affect exchange rates. I now turn to a discussion of suggested ways in which sterilized intervention may possibly affect exchange rates.

The literature on foreign exchange market sterilized intervention posits three such ways:[3] (i) the need for portfolio balance; (ii) announcement effects in the presence of market inefficiencies; (iii) signaling.

(i) The portfolio balance approach assumes that because of foreign exchange risk, foreign and domestic securities are imperfect substitutes, so investors have to be compensated by a risk premium for holding foreign securities. Sterilized intervention changes the supply of foreign relative to domestic assets, thus disturbing equilibrium in investors' portfolios. To restore balance, a change is required in the risk premium measured as the deviation from uncovered interest rate. Hence, the realignment of asset returns produces a change in the spot exchange rate.

---

[2] This section draws heavily on Bordo and Schwartz (1991).

[3] Dominguez (1993) also presents evidence that Federal Reserve sterilized intervention in the 1985–1991 period generally reduced daily exchange rate volatility in accordance with central bank concern to calm disorderly markets.

Empirical evidence, however, generally does not support the assumption that private investors distinguish between foreign and domestic assets (Edison, 1993; Dominguez and Frankel, 1993a). Research in the 1970s tested the joint hypotheses of uncovered interest rate parity—equality of returns between the securities of two countries adjusted for expected change in the exchange rate—and speculative efficiency in the exchange market—the current spot exchange rate is the best predictor of next period's spot rate. The data failed to reject these hypotheses, and early studies concluded in favour of perfect substitution (Mussa, 1979).

Later studies analysed longer data sets with different econometric techniques, leading to the opposite conclusion, namely, actual returns on securities denominated in different currencies differ by more than a random factor. According to these studies, the returns are separated by a time- varying risk premium, determined by relative supplies of domestic and foreign securities (Hodrick, 1987). Supplies of domestic and foreign assets would have to be changed enough to change the risk premium. These findings in turn were challenged by evidence that the risk premium measure is not significantly affected by relative asset supplies in regressions testing for this effect (Frankel, 1982, for risk premia on six currencies and country securities; Rogoff, 1984, for the US and Canadian dollars and their respective bonds). Perfect substitutability was reaffirmed by regressions of the exchange rate on domestic and foreign money supplies, as well as on domestic and foreign assets and wealth, with coefficients on domestic and foreign securities not significantly different from each other.

Recent studies using capital asset pricing models with time-varying risk premia and consumption-based asset pricing models have low explanatory power in validating the portfolio balance justification for sterilized intervention (Obstfeld, 1988). The evidence on risk premia has not been able to account for the deviations from equality of returns on securities of different countries denominated in different currencies and for determining any effects on exchange rates.

Findings generally favourable for the effectiveness of intervention, based on survey data on exchange rate expectations, include statistically significant effects through the portfolio channel and also through an expectations channel (Dominguez and Frankel, 1993b). When an equation for the portfolio channel was estimated, using instrumental variables, perfect substitutability of mark and dollar assets was rejected for the sample period, January 1985-December 1988, implying that Federal Reserve and Bundesbank intervention had a significant although small effect on exchange rates even if sterilized. (Note that no intervention by the Federal Reserve occurred in 1986, and in 1985 and 1987 when it did occur, the amounts involved were small). The effect, however, is heightened if the public has information about the intervention. The expectations channel operates to reduce the ratio of the portfolio demand for marks to the portfolio demand for dollars and thus leads to the depreciation of the mark exchange rate.

© Sc onomic Society 1996

Three possible ways of rationalizing the generally inconclusive results for portfolio balance models have been suggested (Marston, 1988): (a) ex post interest differentials may reflect not risk premia but the failure of traders to take advantage of profit opportunities in the foreign exchange market; (b) risk premia may explain the differentials, but changing asset supplies may have little effect on the exchange rate; and (c) economists possibly lack the technical skills to measure the effects of sterilized intervention on portfolio decisions. None of these suggestions is compelling.

(ii) The second way that sterilized intervention could conceivably affect exchange rates is by conveying information that traders otherwise overlook: they fail to recognize profit opportunities because the foreign exchange market is inefficient. This approach does not explain why traders are less well informed than government bureaucrats. Moreover, traders must guess at the message of intervention when it has taken place, since the authorities choose not to announce intervention actions until some time after the fact. The advantage of this means of conveying information, whatever its content, is obscure.

(iii) Intervention is supposed on this approach to signal a change in the monetary policy intentions of the authorities. The signal activates a response from the private sector to move foreign exchange rates in the direction the authorities seek. How can they be signaling anything if announcements are delayed?

Analysis of intervention signals is murky. One attempt compares the Carter administration's unsuccessful use of sterilized intervention to support the dollar in late 1978 and the Plaza Agreement, which the author interprets as having succeeded in bringing down dollar exchange rates thanks to credible policy announcements (Marston, 1988).

One econometric study reports evidence that Federal Reserve intervention in the 1977–1981 period provided information useful to predict future monetary policy (Dominguez, 1990). The author found a significant relationship from October 1979 to the following spring between money surprises—the difference between weekly money supply growth rates and market forecasts of these growth rates—and the authorities' purchases of foreign currencies between forecast and actual money announcements. Intervention, according to this study, signaled information about money that the earlier forecast did not contain.

Another study found that though intervention in the period 1985–1989 signaled a change in US monetary policy, on average the signals of future monetary policy were opposite in sign to those produced by intervention (Kaminsky and Lewis, 1993).

Obstfeld (1990) suggests that intervention may provide more information than public debt management policies because unanticipated exchange rate changes can have an effect on government's net worth. His example is sterilized

© Scottish Economic Society 1996

purchases of foreign exchange. They signal future depreciation of the currency. However, if the currency appreciates, the government will sustain greater losses than would otherwise have been the case. To avoid this embarrassment, the government's intervention will prove credible to the market even with perfect substitutability of country securities.

Obstfeld's example has a different resonance since he wrote. The Federal Reserve under Alan Greenspan has staked its credibility on achieving zero inflation over a period of years, signaling an appreciating dollar. Yet since 1988 its sterilized purchases of foreign currencies have sent a contrary signal. Dollar depreciation that has accompanied the purchase program, however, has increased the current market value of the currencies acquired. Possible losses on foreign currencies do not therefore appear to provide the kind of information that Obstfeld envisioned. Budgetary effects do not appear to act as a strong constraint on intervention.

In the research on signaling, coordinated intervention is assumed to be more effective than unilateral intervention. It is signaling by several central banks rather than one alone that leads to desired exchange rate changes. A study of daily intervention data for the Bundesbank and the Federal Reserve from 1985 to 1987 shows that coordinated and unilateral sterilized intervention have different and significant effects on the exchange rate premium (Dominguez, 1990). The investigator finds the stronger effects of coordinated than of unilateral intervention to be evidence that market participants interpret policy coordination as a sign of credible commitment. On the other hand, a central bank that participated in a coordinated intervention that signaled a policy change that it did not intend to implement would lose credibility when the foreign exchange market learned of the duplicity.

Three questions may be raised regarding a putative signal that sterilized intervention conveys (Obstfeld, 1990): (a) Is the intervention information an addition to the policy announcement that either accompanies or substitutes for intervention? (b) Why is sterilized foreign exchange intervention more effective as a signal, say, than an open market sale of domestic securities that is sterilized by an increase in central bank loans to commercial banks? (c) If one assumes that signaling is in fact taking place, are the intentions that sterilized intervention signals accepted at face value by the market?

A sceptical answer to these questions reflects the view that central banks need not engage in such a devious way of informing the public of its objectives and the policies it favours to achieve them. Sterilized intervention as a way of signaling or in any other way that has been suggested cannot be shown to change exchange rates.

Determination of exchange rates

From a long-run point of view, depreciation has marked the nominal and real foreign exchange value of the dollar since the end of the Bretton Woods era. Alternating phases of appreciation and depreciation occurred within this overall movement. Two periods of weakness in the mid- and late 1970s—in the latter period the dollar value reached its all-time low—were followed in 1981–1984 by a sharp appreciation of the dollar against the currencies of US trading partners. The entire upward movement was reversed in 1985–1986, and the dollar continued to fall in 1987 to a level not much above that in 1979. In 1989 the dollar appreciated somewhat, followed by renewed weakness in 1990–1992. After strengthening in 1993, the dollar fell in 1994 and the first quarter of 1995, particularly against the mark and the yen.

© Scottish Economic Society 1996

Analysis of the determinants of changes in bilateral exchange rates is simpler than the analysis of changes in the value of the dollar measured on a trade-weighted basis against the currencies of the nine major foreign industrial countries or of the currencies of a larger group including US trade partners among less-developed countries. Identifying the influences on changes in bilateral exchange rates limits the focus to developments in two economies; the broader measures require weights on a composite of influences.

It is also necessary to distinguish trend movements covering decades from swings about the trend operating over a span of several years. There is no dispute that day-to-day fluctuations in foreign exchange markets are noisy, reflecting random influences of 'news' and rumours. However, when it comes to specifying the fundamentals to which long-run movements in exchange rates and swings above and below those movements respond, there is less unanimity. In addition, models of exchange rate determination have been notable for their empirical failure to explain or predict actual exchange rate behaviour.

Here is a catalogue of the forces identified in the literature to account for swings in exchange rate movements:

(i) Monetary policy is one such force. Acceleration of US monetary growth relative to monetary growth in other countries leads to expectations of higher inflation here than abroad and a depreciation of the dollar exchange rate. Deceleration of US monetary growth has the opposite effect. The two episodes of depreciation of the dollar in the 1970s as well as the subsequent ones in later years have been associated with expansionary monetary policy. Nominal dollar depreciation is associated with the excess of the US inflation rate over that of its trading partners. Successful control of inflation alters expectations concerning the future course of inflation and the willingness of authorities to combat it should it again rise. A strong real appreciation will follow. The initial appreciation of the dollar at the start of the 1980s has been associated with contractionary monetary policy (Mussa, 1994). Some investigators, however, do not accept the view that monetary policy is primary in producing medium-term exchange rate movements (Frankel, 1994).

(ii) Movements in the US current account lag movements in the US dollar. Floating exchange rates are not immune to real disturbances from abroad, including not only oil shocks or protectionist measures from abroad but also adverse changes in terms of trade as well as real changes resulting from monetary forces. Real demand for US exports declines as a result of deflation in a partner country.

Persistent deterioration of the trade and current account balance provides information suggesting that a lower real exchange rate is needed to achieve a

© Scottish Economic Society 1996

sustainable balance of payments. Belief that the supply of and demand for foreign exchange adjusts to clear the trade balance, however, ignores the role of financial transactions in foreign exchange markets. Trade deficits did not produce an equilibrating depreciation of the dollar because they were financed by capital inflows from abroad.

(iii) One explanation of the long-term depreciation of the dollar is that years of trade deficits have resulted in the accumulation of a large net foreign investment position in the United States. The flow of foreign capital was essential to finance much of US net investment. When domestic spending exceeds income, a current account deficit must result. The real exchange rate reacts to the current account deficit without a rise in interest rates.

(iv) US fiscal policy is often cited as a determinant of the foreign exchange value of the dollar. Expansionary fiscal policy supposedly pushes up US real interest rates relative to foreign real interest rates, making US goods relatively more expensive than foreign goods. On this view, the real appreciation of the US dollar accounted for the deterioration of the US current account balance in 1980–1984. The increased demand for capital because of the fiscal deficit and high US interest rates produced a massive inflow of foreign capital that was the counterpart of the current account deficit.

(v) Differences in expected real returns to investment and expected productivity growth affect exchange rates. Tax policy that shifts resources from consumption to investment by reducing taxes on saving raises the expected return to capital, thus strengthening the economy and the exchange value of the dollar. Such policy shifts have been cited as contributing to the appreciation of the dollar during 1980–1984.

What conclusion can be reached from the foregoing catalogue of suggested influences on the exchange rate? Clearly structural and macroeconomic factors impinge on the market's view of the value of the US dollar. Over the past two decades, the dollar has been falling on and off. Dollar depreciation has been attributable to a decline in the real dollar, not relatively high US inflation or inflation expectations.

The long-run dollar decline appears to reflect lower relative productivity in the United States, a decline in the US share of worldwide output, and a widening long-run imbalance between US saving and investment, that is, a widening current account deficit. Accompanying US rising reliance on foreign capital inflows is a shrinking relative demand for dollar-denominated assets.

The current episode of dollar weakness has occurred despite strong economic growth, low inflation, contractionary monetary policy and declining real money balances since early 1994. But exchange rates do not move in response to existing and past economic conditions. They move in response to expected future economic conditions.

Will equilibrating forces come to the fore? Dollar depreciation will boost exports. In addition, contractionary monetary policy will reduce consumption and imports by suppressing household purchasing power. The relative price of imports will rise, although exporters from strong currency countries (Germany and Japan) will not be able to raise their prices materially because of weakening product demand. Moreover, exporters from weak currency countries (Mexico and Canada) will provide US consumers with lower- priced imports. If these changes are in the wind, the dollar's exchange value will in time reflect them.

© Sco' onomic Society 1996

Despite heavy intervention by the Bank of Japan and modest intervention in 1995 by the US monetary authorities, the value of the dollar has barely strengthened. Intervention by the Federal Reserve on an unsterilized basis by further hiking the Federal Funds target would temporarily provide support for the dollar. The price, however, would be recession in 1996 with predictable further dollar weakening.

IV WHY DO CENTRAL BANKS INTERVENE?

If assets denominated in major currencies are close substitutes rendering sterilized intervention without effect on exchange rates, what motives drive central banks to intervene? They intervene for at least three reasons: it troubles them that exchange rates (i) are transitorally volatile; (ii) exhibit medium-term variability; and that (iii) those not participating in coordinated intervention might be regarded as displaying a lapse of good citizenship in the world community.

Finance ministers play a role in the decision by central banks to intervene, so political considerations related to foreign policy objectives may be assumed. In the United States the Treasury initiates intervention moves; the timing and size of intervention, however, may be negotiable by the Federal Reserve.

(i) The reason most often cited by the US authorities for intervening is 'disorderly markets.' Financial asset markets are disorderly in the sense of large daily movements above and below the mean daily value. Examples are stock prices, short-term interest rates, long-term bond yields and commodity futures. Exchange rates in fact are less "disorderly" than the prices of other financial assets. If exchange rates mirror unanticipated changes in underlying economic factors, volatility must result.

One interpretation of a high degree of exchange rate volatility is that foreign exchange markets efficiently spread the effects of economic shocks. An unanticipated policy change that leads to portfolio adjustments by investors will change exchange rates and prices of other financial assets before prices of goods move. If asset markets did not quickly absorb the initial shock, prices and quantities of goods and labour markets would have to adjust much faster.

An alternative interpretation of exchange rate volatility is that they are evidence of bandwagon effects, destabilizing speculation, or market inefficiency. Tests of market inefficiency are inconclusive and no criterion has been established for determining that volatility is excessive.

(ii) In addition to concern about disorderly markets that are transitory, central banks show concern about the intermediate-term movements affecting the level of exchange rates. Fluctuations in the level cause changes in other economic variables—aggregate and sectoral output, price level, volume of international trade, and foreign investment flows. A strong dollar stimulates imports and discourages exports. It leads to trade deficits, causes unemployment in export and import-competing industries, and arouses protectionist pressures. A weak

© Scottish Economic Society 1996

dollar raises inflationary concerns as prices of imports rise, and are expected to pass through to nontradable goods prices and to wages.

Thus fluctuations in the level of exchange rates impose adjustment and uncertainty costs on the economy. It is wasteful to shift resources toward and away from tradable goods industries in adjusting to alternate dollar weakening and dollar firming. Uncertainty about the stability of returns on foreign trade and investment also has a resource cost.

(iii) A third reason that central banks intervene is their desire to behave in accordance with expectations that they will participate in programs of coordinated intervention by an international group of countries. Failure to participate, even to a minimal extent, central banks fear, will be interpreted as a dereliction of good standing in the world community. This attitude prevails whether or not countries find coordinated intervention works.

Counter to these central bank motives for intervention, there are reasons to doubt that intervention can achieve the results that are sought.

(i) Large and reversible transitory movements of some exchange rates and widening of bid-ask spreads may be temporarily reduced by intervention. There is no indication, however, that, as often as the US authorities have taken action to quell disorderly markets, any lasting change has been the outcome. Who has benefited from such intervention?

(ii) Whether the dollar is deemed strong or weak, intervention does not deal with the fundamental economic conditions that underlie medium-term changes in the exchange value of the dollar, assuming consensus exists on the specific underlying conditions. The feared consequences of medium-term variations in the dollar's exchange value have not materialized.

The supposedly grossly overvalued dollar in 1981–1985 did not deindustrialize the US economy and usher in a service economy of low-income occupations. Shifts in employment from traded to nontraded goods industries occurred, but such shifts among industries are not uncommon even when the exchange value of the dollar is not involved. Intervention as an answer to protectionist antitrade moves seems a disparate remedy.

Similarly, concern that a weak dollar would induce domestic inflation has not been substantiated. What may be true for small, open economies does not apply to a large, relatively closed economy like the United States. If inflation has at times manifested itself, it has been home-grown by Federal Reserve monetary policy, not an import-driven rise in prices.

In general, the costs of floating exchange rates should be compared to the costs of the alternative, which is fixed or pegged exchange rates, vulnerable to speculative attacks and bigger adjustments.

(iii) The success of coordinated intervention depends less on whether it is undertaken wholeheartedly because countries want to be known as members in good standing than on whether it supports the direction in which the market is moving on its own. Official intervention then would be perceived as committing the authorities to adjust underlying economic policies that influence the behaviour of exchange rates. The evidence that it may be effective as a signaling device is tenuous. It is clear that coordinated intervention is a fair-weather instrument because countries have independent interests at times that they will not sacrifice for the sake of the collectivity. Depreciation of the dollar may be of less concern to other countries than the appreciation of their own currencies.

© Scottish Economic Society 1996

V CONCLUSION

Departing from its scripted passive role under the Bretton Woods system, in 1962 the US began to intervene in foreign exchange markets to defend the dollar. Defending a currency under pegged exchange rates may not appear unreasonable, but why the practice continues under a floating rate system is harder to understand. The swap lines initiated in 1962 have remained in place since and the amount of foreign currencies the lines provide for intervention over the years has tripled. Foreign currencies owned outright by the Federal Reserve, which were trivial in amount before 1979, have increased to a greater extent, especially so in the past decade, with ESF balances somewhat lower than the Federal Reserve's. In any event, the financial resources of private agents in foreign exchange markets are far larger than those of the US, Germany and Japan, and can easily overwhelm official attempts to fix exchange rates.

By accumulating large foreign currency balances the US has made intervention easier. The question is, what has made it more desirable to intervene since the late 1980s than earlier? Holding larger balances does not help the authorities know what exchange rates ought to be. Intervention does not come to grips with underlying monetary and fiscal policies that must change if the objective is to stabilize exchange rates. The propensity of central banks to intervene gives the impression that by doing so they can avoid or at any rate delay adjusting the policies on which markets base their judgments about the exchange value of major currencies.

Official intervention is not an independent tool of economic policy. Exchange rate variability primarily reflects a large range of uncertainty about and instability in basic economic and financial policies. Market judgments about future monetary and fiscal policies and future relative price levels move exchange rates that lead to changes in asset preferences.

One of the indictments of intervention is that it squanders taxpayers' money. An example is the British government's reputed loss of ten billion pounds in September 1992 when it sold foreign currencies in the vain attempt to defend its unrealistic parity against the mark in the Exchange Rate Mechanism. Another example is the Japanese government's reputed loss of tens of billions of dollars in countering the depreciation of the dollar at various times. Other governments are also reputed to have borne intervention losses. These are reputed losses because governments do not routinely inform their publics about the budget consequences of intervention.

The US authorities claim not to have sustained realized losses in the past decade. The record will show gains when the US dollar depreciates and its foreign assets are then worth more, and the record will show losses when the US dollar appreciates and its foreign assets are then worth less. Such results hardly vindicate intervention.

© Scottish Economic Society 1996

The United States for a few years in the 1980s refrained from intervening. It can happen again if its authorities acknowledge that intervening is an exercise in futility.

ACKNOWLEDGEMENTS

Michael D. Bordo made a substantial contribution to this paper. He will be co-author of a monograph on foreign exchange market intervention which he and the present author intend to write.

REFERENCES

BORDO, M. D. (1993). The Bretton Woods international monetary system: a historical overview. In M. D. Bordo and B. Eichengreen, (eds.), *A Retrospective on the Bretton Woods System*. University of Chicago Press, pp. 3–98.
BORDO, M. D. and SCHWARTZ, A. J. (1991). What has foreign exchange market intervention since the Plaza Agreement accomplished? *Open Economies Review*, 2, pp. 39–64.
CATTE, P., GALLI, G. and REBECCHINI, S. (1994). Concerted interventions and the dollar: an analysis of daily data. In P. B. Kenen, F. Papadia and F. Saccomanni (eds.), *The International Monetary System*. Cambridge University Press, pp. 201–239.
DOMINGUEZ, M. K. (1990). Market responses to coordinated central bank intervention. *Carnegie-Rochester Series on Public Policy*, 32, pp. 121–163.
DOMINGUEZ, K. M. and FRANKEL, J. A. (1993a). *Does Foreign Exchange Intervention Work?* Washington, DC: Institute for International Economics.
DOMINGUEZ, K. M. and FRANKEL, J. A. (1993b) Does foreign-exchange intervention matter? the portfolio effect. *American Economic Review*, 83, pp. 1356–1369.
EDISON, H. J. (1993). The effectiveness of central bank intervention: a survey of the post-1982 literature. *Special papers on International Economics*. Princeton University.
FRANKEL, J. A. (1982). In search of the exchange risk premium: a six-currency test assuming mean-variance optimization. *Journal of International Money and Finance*, pp. 255–274.
GREENE, M. L. (1983). US experience with exchange market intervention September 1977–December 1979. Board of Governors of the Federal Reserve System, *Staff Series*, No. 128.
HODRICK, R. J. (1987) The empirical evidence on the efficiency of forward and futures foreign exchange markets. *Fundamentals of Pure and Applied Economics*, 24. Chur, Switzerland: Harwood Academic Publishers.
JURGENSEN, P. (1983). Report of the working group on exchange market intervention [Jurgensen Report]. US Treasury Department.
KAMINSKY, G. L. and LEWIS, K. K. (1993). Does foreign exchange intervention signal future monetary policy? NBER Working Paper No. 4298.
MARSTON, R. C. (1988). Exchange rate policy reconsidered. In M. Feldstein, (ed.), *International Economic Cooperation*. University of Chicago Press, pp. 79–136.
MCKINNON, R. I. (1993). International money in historical perspective. *Journal of Economic Literature*, 31, pp. 1–44.
MUNDELL, R. (1993). Panel session II: implications for international monetary reform. In M. D. Bordo and B. Eichengreen (eds.), *A Retrospective on the Bretton Woods System*. University of Chicago Press, pp. 604–612.
MUSSA, M. (1979). Empirical regularities in the behavior of exchange rates and theories of the foreign exchange market. In *Carnegie-Rochester Conference Series on Public Policy*, 11, pp. 9–57.
OBSTFELD, M. (1990). The effectiveness of foreign exchange intervention: recent experience, 1985–1988. In W. H. Branson, J. A. Frenkel and M. Goldstein (eds.), *International Policy Coordination and Exchange Rate Fluctuations*, University of Chicago Press, pp. 197–237.

© Scottish Economic Society 1996

ROGOFF, K. (1984). On the effects of sterilized intervention: an analysis of weekly data. *Journal of Monetary Economics*, 14, pp. 133–150.
TODD, W. F. (1992). Disorderly markets: the law, history and economics of the exchange stabilization fund and US foreign exchange market intervention. In G. Kaufman, (ed.), *Research in Financial Services: Private and Public Policy*. Greenwich, Ct.: JAI Press, 4, pp. 111–179.
Transcript, Federal Open Market Committee Meeting (1990). 27 March.
TRUMAN, E. M. (1994). Discussion. In P. Kenen *et al.* (eds), *The International Monetary System*, Cambridge University Press, pp. 248–254.
WEBER, A. A. (1994). Foreign exchange intervention and international policy coordination: comparing the G-3 and EMS experience. Centre for Economic Policy Research, Discussion Paper Series No. 1038.

Date of receipt of final manuscript: May 1996.

© Scottish Economic Society 1996

"BEST
ORIGINAL
AVAILABLE"

J. OF EUROPEAN ECONOMIC
V. 22, #2, 1993.

The Role of the Bank for International Settlements in Central Bank Cooperation

Kristin Howell
The University of North Carolina at Wilmington

I

Economists have long debated the potential for and value of cooperation between central banks and the role of such international institutions as the Bank for International Settlements (BIS) in promoting it. The continued interest is due to concerns about the stability of the current exchange-rate structure and the reconciliation of domestic and international policy in an interdependent world.[1] This paper addresses the question of the success of the BIS in fulfilling one of its original functions of promoting central bank cooperation.

The BIS was established in Basle, Switzerland in 1930 with the primary function of transferring World War I reparations payments between Germany and its creditors and the auxiliary function of fostering central bank cooperation.[2] Early cooperative actions involved the buying and selling of domestic and foreign exchange by the central banks and the BIS to reduce currency pressures in the short run. The Bank also used its time deposits from central banks to make short-term investments, often in marketable treasury bills, and to grant credits. The Bank originally possessed the reserves to smooth minor exchange disturbances in a fairly stable monetary environment, but was handicapped in its development by the chaos of the financial crisis of 1931.[3] Central banks were then encouraged to extend credits to other central banks, attempting to coordinate monetary policy and restore confidence in the financial system. Despite weaknesses in these early efforts, the tools for future bank interactions, with the goal of stable economic and monetary growth, were gradually being developed through the BIS.

During the 1930s, politicians and economists discussed the possibilities of a world credit system to help finance growing international trade, the regulation of foreign exchanges, and a gold clearing system. The relationships between central banks and the potential for their cooperation were further considered.

---

[1] Yeager, 'Domestic Stability Versus Exchange Rate Stability', in *Cato Journal*, 8 (1988) pp. 261-77.
[2] Einzig, *World Finance 1914-1935* and *The Bank for International Settlements* and Dulles, *The Bank for International Settlements at Work*.
[3] Einzig, *World Finance 1914-1935*, p. 276.

WHARTON REPROGRAPHICS

6
9

There was debate as to the present value and future scope of activities for BIS. Although not fully coordinated, the foundations for modern central b cooperation developed, partially out of efforts by the BIS to stabilize curren during this often turbulent post-World War I period.

World War II interrupted the progress of the BIS in matters of internati monetary cooperation. The Bank kept a low profile during the war years. keeping with its original position of political neutrality towards individ governments. After the War, problems of post-war economic reconstruct preoccupied the world and many of the negotiations took place at the BIS. question of the role of gold in the international payments system remained well as questions concerning the correct values of the various currenc multilateral payments clearings possibilities, and the value and methods central bank collaboration.

The literature of the time concentrated on discussions of the new Bre Woods system of 1944, a gold-exchange standard, and the Internati Monetary Fund (IMF). The IMF was to oversee the adjustable exchange- system and to give credit to nations with temporary balance of payments def if they were unable to correct these using their international reserves.

What did the future hold for this new system? Would the BIS and the IMF able to work together, or at least to coexist, each with a distinct and us purpose? Or would the BIS be ended as had been suggested under Bre Woods? This last possibility was only briefly entertained and this paper discu the evolution of the BIS into an independent international organization capabl supplementing and expanding the activities of the IMF.[4]

II

Under the Bretton Woods system, exchange rates were linked through 'adjustable peg' system with par values in terms of the dollar or gold. Gold maintained in its position of ultimate reserve asset and the dollar was pegge gold at the 1934 par value of $35 an ounce. Britain's John Maynard Keynes America's Harry Dexter White had proposed at the Bretton Woods Confere that the relatively fixed exchange rate could be periodically adusted wl warranted by a fundamental balance of payments disequilibrium. T condition was broadly defined by the IMF as 'a large actual or potential def or surplus persisting over several years'.[5] Yet nations would prove to unwilling to adjust their rates until exchange speculation became extreme. Th feared loss of national prestige, increased speculation, and the inco redistribution problems of devaluation. Destabilizing international capital flo

---

[4] Schloss, 'The Bank for International Settlements', p. 22.

[5] Salvatore, *International Economics*, p. 536.

especially during the 1960s, were often the result of this hesitation. There seemed to be a contradiction in meaning in the Bretton Woods interpretation that par values should be stable yet adjustable.[6]

The prevailing exchange rate was maintained through official central bank intervention on the exchange markets. Beginning in the late 1940s, governments also frequently used exchange controls to prevent balance of payments imbalances by rationing the available supply of a scarce foreign currency. These controls were meant to impose restrictions on additional capital outflows. The question of sterilizing, or offsetting, undesirable reserve flows through the purchase or sale of domestic securities was also often confronted. These aspects of the Bretton Woods system allowed the BIS role of expanding central bank cooperation to evolve further.

A dollar shortage existed in Europe until 1958 as Europe demanded more imports from the United States. Recovery from World War II proceeded though, and as European trade balances improved, a convertible dollar surplus developed. Confidence in the dollar wavered in the late 1950s as the United States began to lose its ability to convert dollars into gold. The BIS and most central banks, especially the Federal Reserve Bank of New York, discouraged large gold purchases from the United States as uncooperative. Extensive speculation in the exchange markets continued during the 1960s, further hreatening the value of the dollar and the level of the United States gold stock. The banks increased their efforts of coordinated intervention on the exchanges n order to stabilize par values and bolster monetary confidence, many of the contacts and much of the coordination taking place through the BIS. The banks ought and sold foreign exchange to relieve currency pressures and defend the arities against 'hot money' flows. Hot money represented frequent and large lows of short-term capital from one monetary centre to another. It caused nstability in foreign exchange rates and in domestic interest rates and was one f the factors that eventually led to the collapse of the Bretton Woods system.[7]

As gold continued to be drained from the United States during the 1960s, here was a crisis of confidence in the dollar value despite President Kennedy's ledge in 1961 to defend the dollar's value from speculation. The Bretton Woods option of changing the exchange rate had not been exercised and disequilibrium continued, partially because of overexpansion in the United tates. Private gold markets were operating, although they were discouraged by MF policy. Since the London market had re-opened in 1954, attention was rawn to the opportunity for arbitrage and profit by buying gold from the United States and selling it in London. This practice, of course, put an even reater strain on gold reserves and the price of gold.

The central bank Governors met regularly at the BIS to voice their concerns about the gold price and expressed their wishes that the Federal Reserve follo an official policy of exchange intervention. The banks and the BIS insisted th the prevailing passive attitude towards gold flows had to be overcome in ord to restore stability. In 1960, Charles A. Coombs of the Federal Reserve Bank New York was the first representative sent to the BIS meetings by the Unite States.

---

[6] McKinnon, 'Monetary and Exchange Rate Policies for International Financial Stabili-', p. 99.

[7] Salvatore, p. 555.

Thus, while the Bretton Woods system was still being rigidly adhered to i the 1960s, efforts were being made by economists and bankers to negotiat reforms to deal with gold shortages and related exchange tensions. The natur of this 'fixed' exchange system and the problems it created fostered cooperatio between the central banks through the BIS.

III

With the world's monetary system dependent on the dollar under Bretto Woods, new cooperative techniques were considered necessary, techniques th IMF could not completely provide. The structure of the IMF containe 'institutional rigidities,' such as limited discretionary authority and technica expertise, occasionally frustrating the efforts of its executive directors.[8] The BI could act more quickly than the IMF because of its independence fron government control and its smaller size. The fact that its members consisted o industrialized countries with similar economic and political structures an problems increased its potential for consensus.

After World War II, the BIS had gradually rebuilt its resources and resume an active role as banker and forum to the central banks. From 1950 to 1958, th BIS acted as the central clearing agent for the European Payments Unio (EPU), an example of growing central bank cooperation. It organized th smooth monthly settlement of bilateral clearing agreements between the centra banks in order to ensure convertibility and promote trade. The EPU wa replaced in 1958 by the European Monetary Agreement (EMA), which the BI also helps to administer.

A number of European countries were now members of the Organization for European Economic Cooperation, later called the Organization fo Economic Cooperation and Development (OECD). The BIS was, and continues to be, active in collecting, interpreting, and distributing statistical data concerning the financing of the external imbalances of the Group of Ten countries (the United States, England, West Germany, Japan, France, Italy, Canada, the Netherlands, Belgium, and Sweden) to the OECD.

During the 1960s and 1970s, greater political stability was accompanied by increased international interdependence between nations. Financial experts ontinued to consider the traditional role of the BIS, its evolution, and its future cope in a changing environment. The monthly meetings were surrounded by ecrecy and formality, yet they also included more personal and informal atherings of the central bank Governors and their advisors. Sophisticated uestions concerning the needs of bank supervision, financial innovation, oreign exchange system reforms, world liquidity, and the Eurodollar market ere raised. The BIS was frequently involved in cooperative activities to tabilize currencies and reserves, often coming to the rescue of exchanges that ere temporarily under speculative attack. Thus, the role of the Bank in nancial integration was developing through its experience with a series of nancial crises. As a rule, consultations and coordination of bank policies icreased, despite conflicting national goals. During the Johnson presidency, the ontinuity of monetary policies was still evident and the BIS worked with the entral banks to demonstrate their unity in the face of continued dependence on e dollar. Yet, efforts to preserve the Bretton Woods system would eventually rove useless.

During this period, the BIS attempted to contribute to the stability and exibility of the international monetary system through its involvement in the eation of such innovative monetary tools as the swap system, or Reciprocal urrency Agreements (RCA's), the Gold Pool, Roosa Bonds, and an expanded stem of research and information. Cooperation continued to develop between e BIS and central banks through efforts to reform the Bretton Woods system y using these new credit facilities, which are described in the next section.

IV

In 1961, negotiations at the BIS produced a central bank syndicate to ordinate official financial intervention and to control the London gold eculation. This Gold Pool would consist of joint gold sales, through the BIS, y the central banks of the United States, Germany, Switzerland, England, Italy, rance, the Netherlands, and Belgium. Converting the dollar proceeds into gold the central banks was discouraged, encouraging them to participate in the ternational effort to stabilize currencies. The gold sales of the central banks ere able to offset the decrease in the supply of gold from South Africa and eady the official gold price. In 1962, more gold sales were necessary for tervention because of increased speculation. The Pool was characterized by oombs as a short-run 'holding operation in a system full of risks'.[9]

In 1962, a currency swap network (RCA's) was designed through the BIS to tend emergency mutual credit accounts between 15 central banks, increase ternational reserves, and, thus, to defend the par value of the dollar and other

---

[8] Coombs, *The Arena of International Finance*, p. 21.

bid., p. 68.

currencies in the foreign exchange markets. The credits were a type of loan
be reversed at the end of three months at the same rate if not used or extende
They were meant to provide a short-term alternative to international ge
payments. Swap lines were periodically increased to absorb potential speculati
pressures and to avoid exchange controls, representing a cooperative reform
the Bretton Woods system. Although the system is still in place today, the
have been no increases in credit lines since the late 1970s due to the availabil
of other credit sources, such as the IMF. Existing levels seem to be consider
adequate for any necessary short-run, emergency action under the pres
floating exchange-rate regime (see Table I).

The BIS and the banks frequently drew or repaid foreign currenc
through their swap accounts during the 1960s and early 1970s, depending
the adequacy of their official reserve levels or hopes of relieving curre
pressures.[10] The BIS often drew dollars to invest in the Eurodollar market a
in other financial markets in order to promote interest and exchange r
stability and international liquidity. It made Swiss and Belgium francs a
German Marks, acquired through borrowing and exchange, market purchas
available for swap line drawing by the banks. Exchange operations of si
billion dollars between the Federal Reserve Bank, the BIS, and the other cen
banks had been undertaken by 1975 alone. During the 1980s, additio
sources of funds have resulted in less active drawing on the swap system.

The sale of Roosa Bonds by the United States Treasury to foreign cen
banks for longer-term operations represented another new credit facility. Th
bonds were treasury bonds denominated in dollars but with an exchange
guarantee. They were intended to discourage the foreign monetary authori
from converting dollars to gold and draining gold from the U.S. The use
swaps, the Gold Pool, and Roosa Bonds by central banks reflected an attitud
increased willingness to coordinate policies internationally in order to def
the value of the dollar and stabilize exchange markets, an attitude the BIS
promoted for many years.

In 1965, an example of the cooperative efforts of the BIS and a grou
central banks was a 'sterling bear squeeze,' or aggressive buying of sterling,
temporarily succeeded in recovering the value of sterling against expe
devaluation. The Labour Party election in England and the turmoil
followed had threatened the sterling parity. In the next year, rene
speculation prompted the BIS to study the problem of protecting sterling
arrange a large joint credit package through bank and BIS contributi
Continuing pressures forced the need for increased swap lines. The Ban
England borrowed from the BIS group of banks in order to resist specula
until the inevitable devaluation occurred in 1967. Although earlier devalua
might have benefited the British worker in the sale of British goods, the ce

---

[10] Coombs, 'Treasury and Federal Reserve Foreign Exchange Operations', 1962-75,
in more recent years, Pardee, Holmes, and Cross.

anks had cooperated with the BIS to coordinate temporary financial aid in
rder to postpone the parity change. They considered this move towards
tability and confidence in the exchange system beneficial to international
usiness and trade.

The Gold Pool remained active in order to stabilize several cycles of excess
lemand for gold and shaken confidence in the gold value. By the end of the
960s, though, central bank attitudes towards the continuation of the Pool
aried. There had been losses due to the sterling devaluation, yet the banks and
he BIS demonstrated unity until the closure of the Pool in 1968. The
iquidation of the Gold Pool precipitated new speculation and currency
ressures.

Another new financial tool, Special Drawing Rights (SDR's), the new
international money' to be used as reserves, was created in 1967 by an
mendment to the IMF Articles of Agreement.[11] SDR's were distributed to
MF members according to their importance in international trade and entered
s reserves on the books of the IMF. The value of SDR's existed on the basis of
greement between the IMF and its members and they were meant to be used
nly between member central banks in order to settle short-term balance of
ayments disequilibria. Despite the fact that the SDR is a fairly stable store of
alue today, its value will shift with changes in any of the currencies in the
basket' of currencies that now define its value. This unpredictability limits its
volution as a principal reserve asset in the international monetary system.

Thus, the trend in the late 1960s was to attempt to substitute the mutual
redit facilities described here, often through BIS and central-bank cooperation,
or international gold settlements.[12] These facilities were primarily short-term
o medium-term in nature and their use became widespread as the volume of
vorld trade expanded. The Bretton Woods system was being gradually
eformed by the use of these sophisticated financial instruments to address the
eed to protect currency values and to provide new sources of international
iquidity.

V

The Nixon administration damaged this new atmosphere of cooperation, reduced bank communications, and contributed to currency chaos with a policy of 'benign neglect' towards the exchange value of the dollar.[13] In 1969, Nixon severely limited the powers of the Federal Reserve as a representative to the BIS. In 1971, the gold window at the Treasury was closed and the use of the new credit facilities that had been developed were discouraged. The dollar began an inconvertible float, while the use of exchange controls and trade restrictions spread. Confidence in the dollar fell with new waves of speculation in the absence of any official defence of the prevailing dollar value and in the face of the inflationary policies of the United States.[14]

---

[11] Salvatore, *International Economics* and IMF, 'Introduction' to *International Financial Statistics Yearbook*.

[12] Coombs, *Arena of International Finance*, p. 188.

[13] Ibid., p. 203.

The Governors of the central banks, meeting at the BIS, voiced obections to the passive exchange policy of the United States. The banks intervened again through swap drawings, but their actions often proved to be inadequate and belated. By 1973, the Bretton Woods system had collapsed as several currencies floated as a bloc against the dollar, while others floated independently. The adjustable peg exchange-rate system had fixed parity rates for any years, often at unsustainable levels, and the change to a system of market-determined solutions exhibited a certain degree of chaos. Proponents of the credit instruments developed through the BIS under Bretton Woods claim that a more extreme financial crisis was avoided in the transition by the use of these short-term stabilizing tools.

Coombs was able to resume his meetings with the BIS in 1974 and negotiations on exchange-market intervention proceeded in the effort to defend the dollar aganst speculation. The United States borrowed foreign exchange by drawing on its swap agreements with foreign central banks. However, currency instability persisted between the European bloc of currencies and the Dollar bloc in the form of exchange rate reactions to sudden, short-run shocks. The dollar was devalued several times by the end of the 1970s.[15] Gradually, formal bank cooperation began to return though, and some of the credit instruments developed during the 1960s were used again. Countries attempted to improve the coordination of their exchange rate policies through a managed float of the principle exchange rates. Yet, uncertainty as to appropriate exchange-rate levels and adjustment methods persists today, due in part to the divergence of national goals. The BIS continues to provide a forum for the negotiation of these differences.

Many central bankers believed that methods and institutions for cooperation and intervention were necessary under the Bretton Woods system in order to reduce the risks of exchange-rate speculation and provide a more stable international financial structure. They felt the need for exchange-rate surveillance as the scope for the use of new credit instruments as reserve assets increased. The stability, flexibility, and liquidity necessary to enable the monetary system to absorb shocks were, and continue to be, prominent issues concerning the banking sector.

---

[14] Carli, 'International Financial Policies', p. 36.

[15] Coombs, *Arena of International Finance*, p. 236.

VI

The suspension of reparations payments by the Hoover Moratorium in 1931 had not destroyed the cooperation function of the Bank for International Settlements, originally a secondary purpose for its existence. The role of central banking was not well developed during the early years of BIS existence. Thus, the technical aspects of international settlements were the most visible activity of the Bank and its contributions to monetary stability may have been underestimated.[16] The financial crisis of 1931 and the Wars further complicated and guided the evolution of the BIS. Although the power of the Bank never reached the heights some bankers and economists had hoped for, it did continue to broaden its unique position in central bank contact and the creation of innovative monetary tools in order to contribute to global financial stability. Its survival, lengthy existence, expanded membership and activities, and respected status indicate that it is capable of providing certain services that the banking community finds valuable.

Over the past sixty years, the BIS has functioned as a non-political arena for the development of additional banking facilities for international financial cooperation between central banks. It employs approximately three hundred and twenty members of staff from seventeen countries. There are presently twenty-nine central banks around the world who have the right of representation and voting, a small size in comparison to the IMF. The BIS still acts as a clearing house system to assist central banks in making multilateral international payments and in managing their reserves. It remains bound to conform to the monetary policies of its member central banks, but functions as an intermediary between them to oversee negotiations to coordinate their domestic policies and administer the European Monetary System (EMS), established in 1979 by the European Economic Community. The Bank accepts deposits from central banks, buys and sells gold and currencies, buys and sells a variety of marketable securities, and makes advances to central banks and other international institutions. In recent years, the Bank has become increasingly active in making loans, with the cooperation and financial support of central banks, to developing countries, such as Mexico, Brazil, and Nigeria.[17] This credit is often designed to bridge the time during which IMF assistance is being negotiated and to promote exchange market confidence.

The BIS fulfills duties as agent or secretariat to several Groups of Experts (Payments Systems, Computer, Monetary and Economic Data Bank Questions), to the Committee of Governors of the Central Banks of the Member States of the European Economic Community, and to the Board of Governors of the European Monetary Cooperation Fund. The Bank acts as Agent to the private European Currency Unit clearing system, Trustee for

---

[6] Dulles, p. 42.

[7] Bank for International Settlements, *Annual Reports*, 1962-89.

international government loans, and Depositary for the loans issued by the European Coal and Steel Community. It seeks to contribute to the soundness o! the international banking system through its Committee on Bankin; Regulations and Supervisory Practices. The BIS operates as a financial researcl centre for the Group of Ten countries and the Organization of Economi Cooperation and Development. Its Eurocurrency Standing Committee collects analyzes and distributes such information as Euromarket developments an lending risks to the banks in order to help them make financial and economi decisions.

In its Annual Report, the BIS describes its own operations as well as glob economic trends. These characteristics of the report were a factor in the Bank survival during the first years of the Bretton Woods system when som economists called the report 'an admirable argument for the continue existence of the BIS.' [18] The international banking community, as well as mar research and business enterprises, uses this publication for insights into prese and future economic and financial situations and policies.

VII

The motivations of central banks often conflict, yet the Bank f International Settlements remains active as the 'central bankers' bank'. Desp the fact that the virtues of central bank cooperation and the coordination monetary policies are debated, it may be said that the BIS has, at the very lea promoted extensive communication and information exchange betwe interdependent and powerful nations. It has also, in many cases over the yea contributed to international financial stability by helping to create monet tools that have reduced interest and exchange-rate volatility. These actions w designed to promote cooperation and stability in the short run, while longer- reforms could be negotiated.

The role and activities of the BIS evolved in response to its environme where monetary tools were developed to meet the changing needs of the cen banks. The Bank has apparently succeeded in filling a niche and continues blend its traditional, unique functions with an attempt to be flexible and ad to its modern financial environment and to relationships with other institutic Thus, it survived several exchange-rate regimes and developed into international institution that is able to co-exist usefully with such others as IMF and the World Bank.

It is not possible here to answer the question of precisely whose interests served by central bank cooperation and exchange-market intervention, bu examining the context of its occurance, its nature, and its outcomes, effect

---

[18] *The Economist*, "The BIS Report', p. 284.

her international variables may be determined. International effects of BIS d central bank cooperative activities are of interest in view of current concerns er the value and appropriate form of intervention, cooperation, and change-rate regime to achieve monetary stability. In light of ever-increasing terdependencies, these effects may provide lessons concerning the problems d benefits of alternative financial systems and organizations.

Table I

FEDERAL RESERVE RECIPROCAL CURRENCY ARRANGEMENTS, 1962-1970

(millions of dollars)

	1962	1963	1964	1965	1966	1967	1968	1969	1970
ance	50	100	100	100	100	100	700	1000	1000
gland	50	500	500	750	1350	1350	2000	2000	2000
therlands	50	50	100	100	150	150	400	300	300
lgium	50	50	50	100	150	150	225	500	500
nada	250	250	250	250	500	500	1000	1000	1000
itzerland	100	100	150	150	200	250	600	600	600
rmany	50	150	250	250	400	400	1000	1000	1000
ly	150	150	250	450	600	600	750	1000	1250
stria	50	50	50	50	100	100	100	100	200
nmark	0	0	0	0	0	100	100	100	200
eden	0	50	50	50	100	100	250	250	250
an	0	0	150	250	450	450	1000	1000	1000
xico	0	0	0	0	0	130	130	130	130
rway	0	0	0	0	0	100	100	100	200
S*	100	100	150	300	400	1000	1600	1600	1600

wiss franc-dollars and other European currencies-dollars

rce: Charles A. Coombs, «Treasury and Federal Reserve Foreign Exchange Operations», *Federal Reserve Bulletin*, 2-1988

Table I (cont.)

FEDERAL RESERVE RECIPROCAL CURRENCY ARRANGEMENTS, 1971-1979

(millions of dollars)

	1971	1972	1973	1974	1975	1976	1977	1978	1979
nce	1000	1000	2000	2000	2000	2000	2000	2000	2000
gland	2000	2000	2000	3000	3000	3000	3000	3000	3000
therlands	300	300	500	500	500	500	500	500	500
gium	600	600	1000	1000	1000	1000	1000	1000	1000
ada	1000	1000	2000	2000	2000	2000	2000	2000	2000
tzerland	1000	1000	1400	1400	1400	1400	1400	4000	4000
many	1000	1000	2000	2000	2000	2000	2000	6000	6000
y	1250	1250	2000	3000	3000	3000	3000	1000	3000
tria	200	200	250	250	250	250	250	250	250
mark	200	200	250	250	250	250	250	250	250

Sweden	250	250	300	300	300	300	300	300	30
Japan	1000	1000	2000	2000	2000	2000	2000	5000	500
Mexico	130	130	180	180	360	360	360	360	36
Norway	200	200	250	250	250	250	250	250	25
BIS*	1600	1600	1850	1850	1850	1850	1850	1850	185

* Swiss franc-dollars and other European currencies-dollars

Source: Charles A. Coombs, «Treasury an Federal Reserve Foreign Exchange Operations», *Federal Reserve Bullet*
1962-1988.

Table I (cont.)

FEDERAL RESERVE RECIPROCAL CURRENCY ARRANGEMENTS, 1980-1988

(millions of dollars)

	1980	1981	1982	1983	1984	1985	1986	1987	19
France	2000	2000	2000	2000	2000	2000	2000	2000	20
England	3000	3000	3000	3000	3000	3000	3000	3000	30
Netherlands	500	500	500	500	500	500	500	500	5
Belgium	1000	1000	1000	1000	1000	1000	1000	1000	10
Canada	2000	2000	2000	2000	2000	2000	2000	2000	20
Switzerland	4000	4000	4000	4000	4000	4000	4000	4000	4
Germany	6000	6000	6000	6000	6000	6000	6000	6000	6
Italy	3000	3000	3000	3000	3000	3000	3000	3000	3
Austria	250	250	250	250	250	250	250	250	
Denmark	250	250	250	250	250	250	250	250	
Sweden	500	300	300	300	300	300	300	300	
Japan	5000	5000	5000	5000	5000	5000	5000	5000	5
Mexico	700	700	700	700	700	700	700	700	
Norway	250	250	250	250	250	250	250	250	
BIS*	1850	1850	1850	1850	1850	1850	1850	1850	1

* Swiss franc-dollars and other European currencies-dollars

Source: Charles A. Coombs, «Treasury and Federal Reserve Foreign Exchange Operations», *Federal Reserve Bull.*
1962-1988.

REFERENCES

Aubion, R., *The Bank for International Settlements, 1930-1955* (New Jersey: Princeton Univ., Int. Finance Section, Dept. of Economics and Sociology, 1955).

______, "Twenty years of international cooperation in the monetary sphere, 1938-1958', Address given to the Basle Economic and Statistical Society (Basle, 1958).

The Bank for International Settlements, *Annual Reports* (Basle, 1930-88).

The Bank for International Settlements, *The Bank for International Settlements and the Basle meetings* (Basle: The Bank for International Settlements, 1980).

The BIS report', *Economist*, (1940), p. 28.

Buiter, W. H. and Marston, R. C., *International economic policy coordination* (Cambridge: Cambridge Univ. Press, 1985).

arli, G., 'International financial policies', in *NBER summary report: international economic cooperation*, ed. by Feldstein, M. (NBER, 1987).

hacholiades, M., *International monetary theory and policy* (N.Y.: McGraw-Hill Book Co., 1978).

larke, S.V.O., *Central bank cooperation, 1924-31* (N.Y.: Federal Reserve Bank of New York, 1967.

oombs, C.A., *The arena of international finance* (N.Y.: John Wlley & Sons, 1976).

______, "Treasury and Federal Reserve foreign exchange operations. (joint interim report)', *Monthly Review, Federal Reserve Bank of New York* (1962-72).

ooper, R. N., 'Economic interdependence and coordination of economic policies', in *Handbook of international economics. volume II*, ed. by R. W. Jones and P.B. Kenen (Elsevier Science Publishers, 1985), pp. 1195-1233.

urrie, D. and Wren-Lewis, S., 'Evaluating blueprints for the conduct of international macro policy', *AEA Papers and Proceedings*, 79 (1989), pp. 264-9.

ealtry, M. G., Manager and Deputy Head, Monetary and Economic Department, The Bank for international Settlements. Interview (1987).

eGrauwe, P., *Monetary interdependence and international monetary reform* (Saxon House, 1976).

ooley, M. P., 'Capital flight — a response to differences in financial risk', *Int. Monetary Fund Staff Papers*, 35 (1988), pp. 422-36.

ulles, E. L., *The Bank for International Settlements at work* (N. Y.: The Macmillan Co., 1932).

ichengreen, B. and Sachs, J., 'Exchange rates and economic recovery in the 1930s', *NBER Working Paper Series*, # 1498 (1984), pp. 1-37.

inzig, P., *The Bank for International Settlements* (1930).

_____, *World finance, 1914-1935* (Kegan Paul, Trench, Trubner & Co., Ltd., 1935).

_____, "The BIS and multilateral clearing', *The Banker*, 70 (1944), pp. 35-9.

_____, 'Scope for the BIS', *The Banker*, 81 (1947), pp. 189-91.

_____, 'The BIS and eurodollars', *Euromony*, 2 (1970), pp. 58 and 60.

deral Reserve Bank of New York, *Federal Reserve Bulletin*, 1962-1970.

eldstein, M., 'Rethinking international economic coordination', *Oxf. Econ. Pap.*, 40, 1988), pp. 205-219.

enkel, J. A., 'International interdependence and the constraints on macroeconomic policies', *Weltwirtschaftliches Archiv*, 122 (1986), pp. 1-644.

wtrey, R.G., *The art of central banking* (Longmans, Green and Co., 1933).

ernational Monetary Fund, *Balance of Payments Statistics Yearbook* and *International Financial Statistics* (Washington, D.C.: IMF, 1962-76).

Jones, R. W., and Kenen, P. B., eds., *Handbook of international economics - volume 2 - international monetary economics and finance* (N.Y.: Elsevier Science Publishing Company, Inc., 1988.

Kenen, P. B., 'Reforming the international monetary system - Exchange rate management: what role for intervention?', *Amer. Econ. Rev.*, 77 (1987), p. 194-197.

______, 'Global policy optimization and the exchange rate regime', *J of Policy Modeling*, 9 (1987), pp. 19--63.

Kouri, P. J. K. and Porter, M. G., 'International capital flows and portfolio equilibrium', *J. of Pol. Econ.*, 82 (1974), pp. 443-467.

Krueger, A. O., *Exchange rate determination* (Cambridge: Cambridge University Press, 1983).

Laszlo, E., *A strategy for the future: the systems approach to world order* (N.Y.: George Braziller, 1974).

McKibben, W. J. and Sachs, J. D., 'Coordination of monetary and fiscal policies in the OECD', *NBER Working Paper Series*, 1800 (1986).

Salvatore, D., *International economics* (N.Y.: Macmillan Publishing Co., 1987).

Schloss, H. H., *The Bank for International Settlements: an experiment in central bank cooperation* (Amsterdam: North-Holland Publishing Company, 1958).

______ "The Bank for International Settlements', *N.Y University Institute of Finance, Bulletin* # 65-66 (1970).

Schumpeter, J. A., *History of economic analysis* (N.Y.: Oxford Univ. Press, 1954).

Yeager, L. B., 'Domestic stability versus exchange rate stability', *Cato Journal*, 8 (1988), pp. 261-277.

WHARTON/REPROGRAPHICS

ECONOMICS FOCUS

Can the G7 ride again?

This year's economic summit of the Group of Seven big industrial countries starts on June 27th in Lyons. Photo-opportunities aside, quite what purpose the meeting will serve is unclear. Some conclude that it is time to reinvent the G7. It might make better sense to uninvent it

JACQUES CHIRAC, president of France and host of this year's economic summit of the Group of Seven, is said to be keen on a greater role for that grouping of wealthy nations. Apparently he wants it to be more effective in co-ordinating national policies, and to be more ambitious in other ways. This is the sort of thing that French presidents, and hosts of economic summits regardless of nationality, always say. What, if anything, does it mean?

Whether anybody needs this club and its plush gatherings is a question that rolls round every year. For once it deserves more than the customary shrug. Two noted students of the international economy, Fred Bergsten and Randall Henning of the Institute for International Economics in Washington, DC, have made a serious attempt* to say what the G7 is for. Summitry, policy co-ordination and the respective roles of the G-whatevers (see chart) are unfashionable but potentially important topics. The study deserves credit for aiming to revive interest in them, for describing the present arrangements with great clarity—and for its zeal. The authors explain why the G7 has gone wrong, and what to do about it.

To see what summits at their best can achieve, they say, look to past successes. The Bonn summit of 1978 adopted a "co-ordinated global growth strategy with precise commitments for action ..." The apogee of co-operation, however, was 1985-87, when the G7 "played a central role in correcting the enormous overvaluation of the dollar ..." Other successes include responses to the debt problems of the 1980s, arrangements for sharing the cost of the Gulf war in 1991, and so on.

In the past few years, the argument goes on, the G7's role has dwindled. This decline is not, as you might think, because the group's weight in the world economy has diminished. The G7's share of world output now is roughly what it was in 1960—about half at PPP exchange rates.

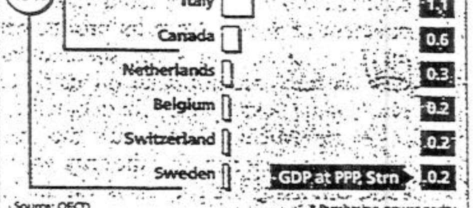
The real reason, according to the study, is partly that old frictions have come more to the fore: America's fondness for fiscal expansion, Germany's for frugality; America's interest in bilateral, not overall, trade imbalances; the greater weight attached to regional co-operation. And partly it is due to new factors, including the consensus which says that "globalisation" has removed the power of governments to steer their economies.

On this last point, the study is surely right to challenge the conventional view. Despite the growth of capital flows, governments retain discretion in economic policy. Policy co-ordination is possible—the question is whether it is desirable, and whether the G7 is the right forum.

Rewriting history

Here, the study is less convincing. Yes, it cites various past G7 successes—but hang on a minute. Most economists regard the summit of 1978 as a failure. Many blame the global inflation that followed, and the subsequent recession, on that very initiative. The exchange-rate co-operation of 1985-87 was hardly an unqualified success. You could argue (many do) that the exercise was bungled, and that the result was the global stockmarket crash of 1987. In countries such as Britain, the crash contributed to errors in monetary policy, and hence to an aggravated cycle of boom and bust.

The main job that the authors want a refurbished G7 to take on (as those who know Mr Bergsten's work will not be surprised to learn) is a renewed attempt to manage exchange rates. But whether exchange-rate targeting makes sense in the first place matters far more than whether the G7 (an unlikely candidate, as it happens) should have a role in it. Here the study gets the cart before the horse.



Source: OECD

*Purchasing-power parity

Note: Finance ministers of the G10 (which actually has 11 members; don't ask why) first met in 1962 to arrange new lending to the IMF. The group still meets, but was replaced as the top forum for economic policy in 1973 by the G5. This was a breakthrough because the G5 has only five members. From 1986 G5 summits became G7 summits. However, the G7 is now really the G8, because the European Commission and Council of Ministers are also represented. If Russia joined, the G7 would presumably become the G8, but really ought to be called the G9. And so on.

Wherever you look, in fact, the record of co-operation is mixed. The debt crisis is no exception. Again, many would question whether it was well handled—and of those who think it was, many would give the credit to America (for its Baker and Brady plans), or to the IMF and the World Bank, rather than to the G7.

There is a deeper problem with the view that the world economy needs the G7 to exercise "leadership". Formal regimes of co-operation in economic policy tend to require that governments are willing and able, now and then, to put group interests ahead of national ones. None of the G7 countries, least of all America, is willing in principle or able in practice to do so. In terms familiar in Britain, none has any intention of surrendering sovereignty to the G7.

The corollary is that the G7 cannot be more than a forum within which to discuss and, above all, present policies that governments would have wished to undertake, for good or ill, anyway. Occasions when that kind of co-operation seems useful may not happen very often. They will typically take the form of an event requiring swift joint action of a clearly discernible nature. Possibly the G7 has seemed less effective of late because there have been fewer such occasions. (The authors ask why the G7's response to the Mexican peso crisis was muddled. The answer is that the right response was by no means clear, and that America's interests were not fully shared by its G7 partners.)

Seen in this light, admittedly, it may be wrong to blame co-operation, as such, for past errors such as the fiscal expansion of the late 1970s. If co-operation is merely a veil, blame instead the mistaken policies of particular governments. But it also follows that it is not in the G7's nature to lead. Its initiatives are bound to be occasional and ad hoc. Messrs Bergsten and Henning deplore this messiness. They seek a neater system, a bold new institution. They seek in vain.

---

* "Global Economic Leadership and the Group of Seven." Institute for International Economics.

WHARTON REPROGRAPHICS

Exchange rates

The fixed and the flexible

The Czech Republic is the latest emerging economy to allow its currency to float within a pre-set band. Why are these bands becoming so popular?

IF SURPRISE is the secret of successful central banking, the Czech National Bank seems to be getting the knack. At the end of February it widened the range in which the Czech currency, the koruna, floats against a weighted average of the D-mark and the American dollar from 0.5% to 7.5% either side of a central value. Some adjustment of the five-year-old limits had been expected; but its timing and extent had not. The Czechs have dumped what was, for all practical purposes, a fixed exchange-rate regime for an increasingly common system among emerging economies: allowing the exchange rate to float within a pre-ordained band.

Exchange-rate bands come in two types. Under the first, the bands are fixed. They are described either by variations around a central rate, as in the Czech system, or by outer limits, as, for example, in Russia. The Russian central bank is committed to keeping the rouble's value between 4,550 and 5,150 to the dollar until June 1st. The second, used by Poland, Chile, Colombia and Israel, is a "crawling band". Here, the central rate (or, equivalently, the band limits) moves according to specific rules. In Israel, the central rate of the shekel against a basket of currencies has depreciated by 6% a year since July 1993. The Chilean peso's course is set monthly.

Currency bands are becoming popular because they seem to offer monetary authorities the advantages of both fixed and floating exchange-rate systems. A commitment to a fixed rate can be a powerful weapon in the fight against inflation. If monetary policy is too loose, then the country's foreign-exchange reserves will automatically fall as investors shift into foreign currency. Defending the currency's fixed exchange rate forces the central bank to tighten its policy.

But there are also disadvantages in this method. If a country attracts a lot of foreign capital, as the Czech Republic has, this foreign money may tend to fuel domestic inflation unless the central bank "sterilises" it by selling bonds, an operation which becomes ever more expensive.

A floating exchange rate, by contrast, confers greater flexibility. If a country is hit by a sudden outside shock, such as a big capital inflow, this can be absorbed by an appreciation of the currency, rather than by domestic inflation. And an outflow of funds will lead to currency depreciation, rather than to painful reductions in nominal wages.

Because taming runaway inflation was a top priority when countries in Latin America and Eastern Europe embarked on their economic reforms, many of them began with fixed exchange-rate regimes. Two factors explain their current appetite for greater flexibility.

Bend a little

First, governments have realised that their economies are vulnerable to rapid reversals of capital flows. The Czechs, for example, have struggled to cope with big inflows; some economists argue that a currency band can reduce these by increasing uncertainty about exchange-rate movements. And the dramatic after-effects of Mexico's financial crisis reminded governments of how quickly foreign money can leave. In both cases a system that allows the exchange rate to vary a little gives a country more room for manoeuvre.

The second reason for the switch to bands is that these can help a country to conquer inflation gradually, without pricing its goods out of international markets in the process. Although they have beaten hyperinflation, inflation rates in many emerging economies are still stubbornly high by the standards of richer countries. With fixed exchange rates, their currencies would gradually appreciate in real terms. But a band, especially one of the crawling variety, allows a country to act against this through a controlled depreciation of its currency's exchange rate.

How successfully governments can use exchange-rate bands to combat inflation without suffering from an overvalued currency depends on the way bands are structured. In Israel, the rate at which the central rate of the shekel is adjusted is based on the difference between the country's inflation target for the year and expected world inflation. This forward-looking approach tries to influence people's expectations about future inflation. In Chile, by contrast, the rate of crawl is based on the previous month's inflation rate and a forecast of world inflation. Since it accommodates past price increases, Chile's system is a slower means of reducing inflation than Israel's, but is less likely to result in overvaluation.

A currency band may make it easier to combine anti-inflationary discipline and exchange-rate flexibility. But it does not remove the trade-off between the two, and it will work only if the government's other policies are consistent with it.

Mexico's experience is a case in point. It had a crawling band before its financial crash in December 1994. The band was used primarily as an anchor against inflation. But Mexico's monetary policies were inconsistent with this goal, the peso became overvalued, and collapse ensued. They may promise the discipline of fixed rates and the flexibility of floating ones, but currency bands are still no substitute for sensible monetary policies.

WHARTON REPROGRAPHICS

13

The Evolution of Modern Global Competition

Alfred D. Chandler, Jr.

THE MULTINATIONAL AND THE TRANSFORMATION OF INTERNATIONAL TRADE

International trade and with it competition for distant markets has, of course, existed for centuries. But the global competition on which this colloquium concentrates is entirely modern. Until the last decades of the nineteenth century, international competitors were merchants and traders of different nationalities who purchased their goods from manufacturers and often from other intermediaries, usually wholesalers. Over the past century, international trade has been increasingly carried on by manufacturing enterprises that marketed their goods through their own salaried sales force operating from their own office in foreign lands. Many of these enterprises established their own manufacturing establishments abroad. Still others set up or acquired integrated foreign subsidiaries with personnel and facilities for marketing, purchasing, and R & D as well as for production.

The new multinationals not only transformed international trade but also the relationship between private business enterprise and public institutions. Trade across borders coordinated by a single administrative hierarchy carrying different economic functions and often handling several different product lines could not be regulated by those traditional means that had regulated market-coordinated flows between merchants and manufacturers of different nations. Moreover, the economic power of the multinationals continued to increase. Changing technologies and expanding markets enhanced the advantages of managerial coordination and at the same time intensified the competition between such managerial enterprises. As the technologies of production and distribution and the markets for products became more homogeneous, such competition forced the multinationals to acquire a broader, more global perspective.

An understanding of the evolution of modern global competition requires, therefore, a description and analysis of the evolution of this new species of business enterprise, the new player in international trade—the integrated, diversified multinational. This chapter attempts to provide such historical background by reviewing why the modern multinational appeared when it did, became located in the industries that it did, and grew to large size by the processes that it did. Once the similarities in the beginnings and early growth of the new institution are examined, the review considers, even more briefly, the significant differences that occurred in this evolutionary development in four of the world's leading industrial nations—the United States, the United Kingdom, Germany, and Japan. Next it examines the ways these similarities and differences shaped global competition in the period between the two World Wars. Finally it outlines basic factors that have transformed the nature of that competition in the years following the World War II.

THE BEGINNINGS OF THE MODERN MULTINATIONAL INDUSTRIAL ENTERPRISE

Let us begin by stressing once again that the large multinational administered by a hierarchy of salaried managers is an entirely modern phenomenon. Until the last decades of the nineteenth century, production and distribution of goods was carried on by small, personally managed enterprises. The rapid expansion of trade and markets in the eighteenth and early nineteenth centuries brought specialization in both the production and distribution. By the mid-nineteenth century, business enterprise normally operated a single unit of production (a farm, mine, mill, or shop) or a single unit of distribution. The latter normally handled a single function (wholesaling, retailing, importing, or exporting) and usually handled a single line of products (textiles, apparel, hardware, furniture, tobacco, drugs, jewelry, and the like). The flow of goods between these enterprises was coordinated by market mechanisms, which in turn were often affected by tariffs, subsidies, and other regulations decreed by nation states. At that time, owners, either as individuals or partners, managed their enterprises. Where salaried managers were employed, they were few in number and worked directly with the owners. As yet there were no middle managers who supervised junior managers and reported to senior ones. Nor were there enterprises administered through hierarchies of managers comparable to that depicted on Figure 13.1.

Technological innovation transformed such personally managed, single-unit, single-function enterprises into multiunit, multifunctional

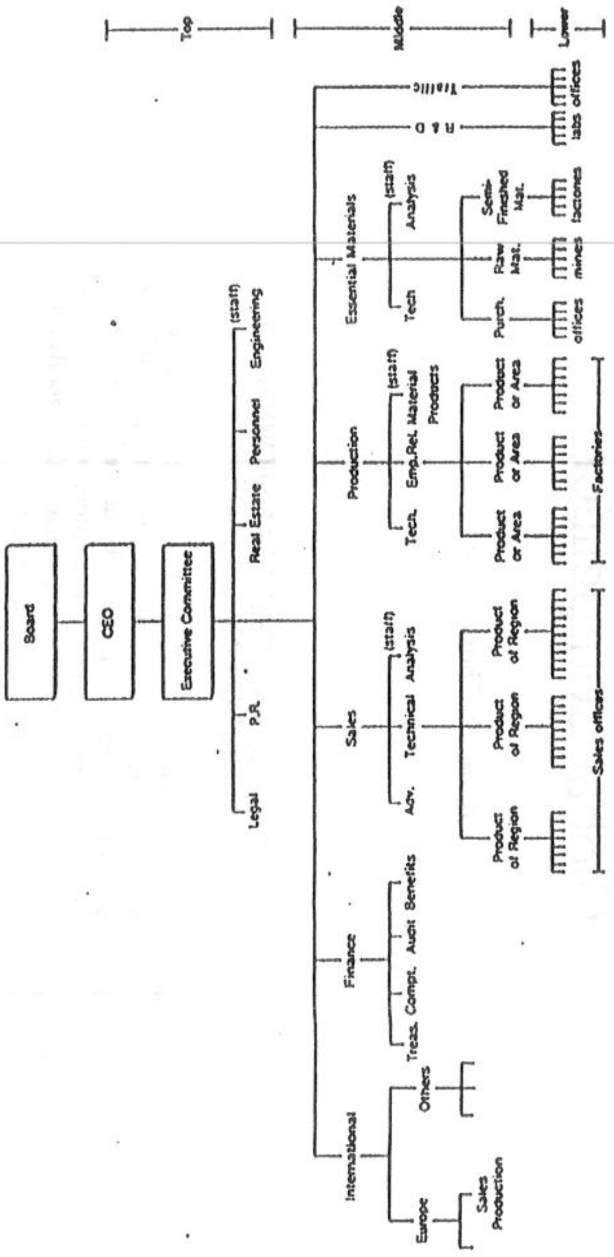

Top
Middle
Lower
Board
CEO
Executive Committee
Legal
P.R.
Real Estate
Personnel
Engineering
(staff)
International
Finance
Sales
Production
Essential Materials
Traffic
R & D
Europe
Others
Sales
Production
Treas.
Compt.
Audit
Benefits
Adv.
Technical
Analysis
(staff)
Tech.
Eng.Rel. Material
Products
(staff)
Tech
Analysis
(staff)
Product of Region
Product of Region
Product of Region
Product or Area
Product or Area
Product or Area
Purch.
Raw Mat.
Semi-Finished Mat.
Sales offices
Factories
offices
mines
factories
labs
offices

Figure 13.1 Multiunit, Multifunctional Enterprise

ones administered through a hierarchy of salaried managers. New technologies of production and distribution made possible historically unprecedented increases in the volume of output that a single unit of production could process and the volume of transactions that a single unit of distribution could handle. The resulting enormous expansion in the flow of goods in domestic and international trade was in large part not the result of the first industrial revolution that began in Britain at the end of the eighteenth century; that is, it was not entirely the result of the initial application of the new sources of energy (fossil fuel, coal) to the processes of production. It resulted much more from the coming of modern transportation and communication. The railroad, telegraph, steamship, and cable made possible the modern mass production and distribution that was the hallmark of the second industrial revolution of the late nineteenth and early twentieth centuries. These new high-volume technologies could not be effectively exploited unless the massive flows of materials were guided through the process of both production and distribution by teams of salaried managers.

Thus, the first such managerial hierarchies appeared during the 1850s and 1860s to coordinate the movements of trains and the flow of goods over the new railroad networks and messages over the new telegraph system. The management of the new mass retailing establishments—the department stores, mail order houses, and chains or multiple shops—whose existence the railroad and telegraph made possible, also required the creation of teams of salaried managers. Such a managerial organization permitted Sears Roebuck's mail order store in Chicago by 1905 to fill 100,000 orders in a single day—more than the average earlier U.S. merchant filled in a lifetime.[1] These administrative hierarchies grew to a still much greater size in industrial enterprises that integrated mass production and mass distribution within a single business enterprise, again on the basis of modern transportation and communication.

The integrated industrial enterprises have had much in common whether they were U.S., European, or Japanese. They appeared in the 1880s and 1890s in the United States and Europe and a little later in Japan, only because Japan was later to industrialize. They clustered in much the same types of industries; and, finally, they grew in much the same manner. In nearly all cases they became large, first, by integrating forward—that is, investing in marketing and distribution facilities and personnel—and by moving backward into purchasing and control of raw and semifinished materials; then, though less often, by investing in research and development. In this way they created the multifunctional organization that is depicted in Figure 13.1. Many of these new enterprises began in the 1880s and 1890s to make direct investments abroad, first in marketing and then in production facilities and personnel. Such enterprises were the first modern industrial

multinationals. Many of these pioneers are, a century later, still major players in today's global competition.

Although not all such integrated industrial enterprises became multinationals, nearly all industrial multinationals evolved from such enterprises. From their beginnings in the 1880s and 1890s, these integrated managerial enterprises have clustered in industries with high-volume processes of production whose output was distributed to national and international markets. Tables 13.1 to 13.5 illustrate this point. Table 13.1 indicates the location by country and by industries of all industrial corporations in the world that in 1973 employed more than 20,000 workers. (The industries are those defined as two-digit industrial groups by the U.S. Census's Standard Industrial Classification [SIC]). In 1973, 263 (65 percent) of the 401 companies were clustered in food, chemicals, oil, machinery, and primary metals. Only 21 companies (5.5 percent) were in apparel, lumber, furniture, leather, publishing and printing, instruments, and miscellaneous. Just under 30 percent were in three-digit subcategories of the other of the census's two-digit groups, such as cigarettes in tobacco; tires in rubber; newsprint in paper; plate glass in stone, glass, and clay; cans and razor blades in fabricated metals; and mass-produced cameras in instruments.

A second point that Table 13.1 illustrates is the predominance of U.S. firms among the world's largest industrial corporations. Of the total of 401 companies employing more than 20,000 persons, over half (212, or 52.6 percent) were U.S. The United Kingdom followed with 50 (12.5 percent), Germany with 29 (7.2 percent), Japan with 28, (7.0 percent) and France with 24 (6.0 percent). Only in chemicals, metals, and electrical machinery were there as many as 4 or 5 more non-U.S. than U.S. firms. In the other groups, U.S. firms total as many or more than those of all other nations.

Table 13.2 shows that in the United States the large industrial corporations had clustered throughout the twentieth century in the same industries in which they were concentrated in 1973. The pattern, depicted on Tables 13.3, 13.4, and 13.5, is much the same for Britain, Germany, and Japan. The U.S. firms were larger, as well as more numerous, than those in other countries. For example in 1948, only 70 to 80 of the British firms had assets comparable to those of the top 200 in the United States. In 1930, the number was about the same. For Germany and Japan it was smaller. Well before World War II, the United States had many more and many larger managerial hierarchies than did other nations—underlining the fact that managerial capitalism first emerged in that nation.

An Explanation of the Beginnings

Why have these large, integrated, hierarchical, and increasingly multinational enterprises appeared in some industries but rarely in oth-

Table 13.1 The Distribution of the Largest Industrial Enterprises with More Than 20,000 Employees, by Industry and Nationality, 1973

SIC		*U.S.*	*Outside of the U.S.*	*U.K.*	*Germany*	*Japan*	*France*	*Others*	*Grand Total*
20	Food	22	17	13	0	1	1	2	39
21	Tobacco	3	4	3	1	0	0	0	7
22	Textiles	7	6	3	0	2	1	0	13
23	Apparel	6	0	0	0	0	0	0	6
24	Lumber	4	2	0	0	0	0	2	6
25	Furniture	0	0	0	0	0	0	0	0
26	Paper	7	3	3	0	0	0	0	10
27	Printing and Publishing	0	0	0	0	0	0	0	0
28	Chemical	24	28	4	5	3	6	10	52
29	Petroleum	14	12	2	0	0	2	8	26
30	Rubber	5	5	1	1	1	1	1	10
31	Leather	2	0	0	0	0	0	0	2
32	Stone, Clay, and Glass	7	8	3	0	0	3	2	15
33	Primary Metal	13	35	2	9	5	4	15	48
34	Fabricated Metal	8	6	5	1	0	0	0	14
35	Machinery	22	12	2	3	2	0	5	34
36	Electrical Machinery	20	25	4	5	7	2	7	45
37	Transportation Equipment	22	23	3	3	7	4	6	45
38	Instruments	4	1	0	0	0	0	0	5
39	Miscellaneous	2	0	0	0	0	0	0	2
	Diversified/ Conglomerate	19	3	2	1	0	0	0	22
	Total	211	190	50	29	28	24	59	401

NOTE: In 1970 the one hundred largest industrials accounted for more than one-third of net manufacturing output in the United States and over 45 percent in the United Kingdom. In 1930 they accounted for about 25 percent of total net output in both countries.
SOURCE: *Fortune*, May 1974 and August 1974.

ers? Why did they appear at almost the same historical moment in the United States and Europe? Why did these industrial enterprises in advanced economies grow in the same manner, first by integrating forward into volume distribution, then taking on other functions, and finally becoming multinational by investing abroad first in marketing and then in production?

Because these enterprises initially grew larger by integrating mass production with volume distribution, answers to these critical questions require a careful look at both these processes. Mass production is an attribute of specific technologies. In some industries the primary way to increase output was to add more workers and machines; in others it was by improving and rearranging the inputs, by improving the machinery, furnaces, stills, and other equipment, by reorienting the process of production within the plant, by placing the several intermediate processes of production required for a finished product within a single works, and by increasing the application of energy (particularly fossil fuel energy). The first set of industries remained "labor-intensive;" the second set became "capital-intensive." In this second set of industries the technology of production permitted much larger economies of scale than were possible in the first. That is, it permitted much greater reduction in cost per unit of output as volume increased. So in these capital-intensive industries with large batch or continuous process production technologies, big factories or plants, operating at minimum efficient scale (scale of operation that brought the lowest unit costs) had a much greater cost advantage over small works than was true with labor-intensive technologies. Similarly costs per unit

Table 13.2 The Distribution of the 200 Largest Industrial Enterprises in the United States, by Industry[a]

SIC		*1917*	*1930*	*1948*	*1973*
20	Food	30	31	26	22
21	Tobacco	6	5	5	3
22	Textiles	5	3	6	3
23	Apparel	3	0	0	0
24	Lumber	3	4	1	4
25	Furniture	0	1	1	0
26	Paper	5	7	6	9
27	Printing and Publishing	2	3	2	1
28	Chemical	20	18	24	27
29	Petroleum	22	26	24	22
30	Rubber	5	5	5	5
31	Leather	4	2	2	0
32	Stone, Clay, and Glass	5	9	5	7
33	Primary Metal	29	25	24	19
34	Fabricated Metal	8	10	7	5
35	Machinery	20	22	24	17
36	Electrical Machinery	5	5	8	13
37	Transportation Equipment	26	21	26	19
38	Instruments	1	2	3	4
39	Miscellaneous	1	1	1	1
	Diversified/ Conglomerate	0	0	0	19
	Total	200	200	200	200

[a] Ranked by assets.

Table 13.3 The Distribution of the 200 Largest Industrial Enterprises in the United Kingdom, by Industry[a]

SIC		1919	1930	1948	1973
20	Food	63	64	52	33
21	Tobacco	3	4	6	4
22	Textiles	26	24	17	10
23	Apparel	1	3	2	0
24	Lumber	0	0	0	2
25	Furniture	0	0	0	0
26	Paper	4	5	6	7
27	Printing and Publishing	5	10	6	7
28	Chemical	11	9	19	21
29	Petroleum	3	3	3	8
30	Rubber	3	3	2	6
31	Leather	0	0	1	3
32	Stone, Clay, and Glass	2	6	7	16
33	Primary Metal	35	18	24	14
34	Fabricated Metal	2	7	9	7
35	Machinery	8	7	10	26
36	Electrical Machinery	11	18	12	14
37	Transportation Equipment	20	14	20	16
38	Instruments	0	1	1	3
39	Miscellaneous	3	4	3	1
	Diversified/ Conglomerate	0	0	0	2
	Total	200	200	200	200

[a] Ranked by sales for 1973 and by market value of quoted capital for the other years.

rose much more rapidly when volume of production fell below minimum efficient scale (of say 80 to 90 percent of rated capacity) than was true in labor-intensive industries.

Table 13.4 The Distribution of the 200 Largest Industrial Enterprises in Germany, by Industry[a]

SIC		1913	1928	1953	1973
20	Food	23	28	23	24
21	Tobacco	1	0	0	6
22	Textiles	13	15	19	4
23	Apparel	0	0	0	0
24	Lumber	1	1	2	0
25	Furniture	0	0	0	0
26	Paper	1	2	3	2
27	Printing and Publishing	0	1	0	6
28	Chemical	26	27	32	30
29	Petroleum	5	5	3	8
30	Rubber	1	1	3	3
31	Leather	2	3	2	1
32	Stone, Clay, and Glass	10	9	9	15
33	Primary Metal	49	47	45	19
34	Fabricated Metal	8	7	8	14
35	Machinery	21	19	19	29
36	Electrical Machinery	18	16	13	21
37	Transportation Equipment	19	16	14	14
38	Instruments	1	2	4	2
39	Miscellaneous	1	1	1	1
	Diversified/Conglomerate	0	0	0	1
	Total	200	200	200	200

[a] Ranked by sales for 1973 and by assets for the other three years.

What is of basic importance for an understanding of the coming of the modern managerial industrial enterprise is that the cost advantage of the larger plants cannot be fully realized unless a constant flow of materials through the plant or factory is maintained to assure effective capacity utilization. The decisive figure in determining costs and profits is, then, not rated capacity for a specified time period but rather throughput, that is, the amount actually processed in that time period. Throughput relative to rated capacity is the proper economic measure of capacity utilization. In the capital-intensive industries the throughput needed to maintain minimum efficient scale requires not only careful coordination of flow through the processes of production but also the flow of inputs from the suppliers and the flow of outputs to the retailers and final consumers. Such coordination cannot happen automatically. It demands the constant attention of a managerial team, or hierarchy. Thus, scale is only a technological characteristic. The economies of scale, measured by throughput, require organizational input. Such economies depend on knowledge, skills, and teamwork—on the human organization essential to exploit the potential of technological processes.

Explanatory Generalizations Illustrated

A well-known example illustrates these generalizations. In 1882 the Standard Oil "alliance" formed the Standard Oil Trust.[2] The purpose was not to obtain control over the industry's output. That alliance, a loose federation of forty companies each with its own legal and administrative identity but tied to John D. Rockefeller's Standard Oil Company through interchange of stock and other financial devices, already controlled close to 90 percent of the U.S. output of kerosene. Instead the Trust was formed to provide a legal instrument to rationalize the industry and to exploit more fully economies of scale. The Trust provided the essential legal means to create a corporate or central office that could, first, reorganize the processes of production by shutting down some refineries, reshaping others, and building new ones,

Table 13.5 The Distribution of the 200 Largest Industrial Enterprises in Japan, by Industry*

SIC		*1918*	*1930*	*1954*	*1973*
20	Food	31	30	26	18
21	Tobacco	1	1	0	0
22	Textiles	54	62	23	11
23	Apparel	2	2	1	0
24	Lumber	3	1	0	1
25	Furniture	0	0	0	0
26	Paper	12	6	12	10
27	Printing and Publishing	1	1	0	2
28	Chemical	23	22	38	34
29	Petroleum	6	5	11	13
30	Rubber	0	1	1	5
31	Leather	4	1	0	0
32	Stone, Clay, and Glass	16	14	8	14
33	Primary Metal	21	22	28	27
34	Fabricated Metal	4	3	6	5
35	Machinery	4	4	10	16
36	Electrical Machinery	7	12	15	18
37	Transportation Equipment	9	11	18	20
38	Instruments	1	1	3	5
39	Miscellaneous	1	1	0	1
	Diversified/Conglomerate	0	0	0	0
	Total	200	200	200	200

*Ranked by assets.

and second, coordinate the flow of materials, not only through the several refineries, but from the oil fields to the refineries and from the refineries to the consumers. The resulting rationalization made it possible to concentrate close to a quarter of the world's production of kerosene in three refineries, each with an average daily charging capacity of 6,500 barrels with two-thirds of their product going to overseas markets. (At this time the refined petroleum products were by far the nation's largest nonagricultural export.) Imagine the diseconomies of scale (that is, the great increase in unit costs) that would result from placing close to one-fourth of the world's production of shoes, textiles, or lumber into three factories or mills!

This reorganization of the Trust's refining facilities brought a sharp reduction in its average cost of producing a gallon of kerosene. It dropped from 1.5 cents a gallon before reorganization to 0.54 cents in 1884 and 0.45 in 1885—costs far below those of any competitor. However, maintaining this cost advantage required these large refineries to have a continuing daily throughput of from 5,000 to 6,500 barrels, or a three- to four-fold increase over earlier 1,500 to 2,000 barrels daily flow. This required concomitant increases in transactions handled and in the complexity of coordinating the flow of materials through the process of production and distribution.

The Trust quickly recruited the managerial team essential to coordinate the flow through the new refineries. Then in the 1880s, it built a marketing network that quickly covered the world and so became one of the very first modern giant industrial multinationals. As late as 1910, Standard Oil still enjoyed 75 percent of the European markets for kerosene and lubricating oil despite the efforts of the Nobels, the Rothschilds, and such enterprises as Royal Dutch and Shell to reduce Standard's share in that rich market.

The Standard Oil story was by no means unique. In the 1880s and 1890s new mass production technologies—those of the second industrial revolution—brought a sharp reduction in costs as plants reached minimum efficient scale. In many industries the level of output was so high at that scale that a small number of plants were able to meet existing national and even global demand. The structure of these industries quickly became oligopolistic. Their few large enterprises competed worldwide. Many of the first enterprises to build a plant with a high minimum efficient scale and to recruit the essential management team have remained until this day leaders in their industries. A brief review of the industries listed in Tables 13.1 to 13.5 in which the large enterprises have always clustered illustrates this close relationship between scale economies, the size of enterprise, and industrial concentration.

In groups 20 and 21—food, drink, and tobacco—the adoption of brand new production processes brought rapid reduction of costs in the refining of sugar and vegetable oils, in the milling of wheat and oats, and in the making of cigarettes. In cigarettes, for example, the invention of the Bonsak machine in the early 1880s permitted the first entrepreneurs to adopt the machine—James B. Duke in the United States and the Wills brothers in Britain—to reduce labor costs sharply, in Wills's case from 4 shillings per 1,000 to 0.3 pence per thousand.[3] Understandably Duke and Wills soon dominated and then divided the world market. In addition most companies in group 20, and also those producing consumer chemicals, such as soap, cosmetics, paints, and pills, pioneered in the use of new high-volume techniques for packaging their products in small units that could be placed directly on retailers' shelves. The most important of these was the "automatic-line" canning process invented in the mid-1880s, which permitted the filling of 4,000 cans an hour. The names of these pioneers, Campbell Soup, Heinz, Bordens, Carnation, Nestlé, Cadbury, Cross and Blackwell, Lever, Procter & Gamble, Colgate, and others are still well known today.

In chemicals, group 29, the new technologies brought even sharper cost reductions in industrial than in packaged consumer products. Large-scale production of synthetic dyes and synthetic alkalis began in the 1880s. It came a little later in synthetic nitrates, synthetic fibers, plastics, and film. The first three firms to produce the new synthetic blue dye, alizarin, dropped production costs from 200 Reich-marks per kilo in the 1870s to 9 Reich-marks by 1886; and those three firms—Bayer, BASF, and Hoechst—are still, a century later, the three largest German chemical companies.[4]

Rubber production, group 30, like oil, benefited from scale economies, even more in the production of tires than rubber footwear and clothing. Of the ten rubber companies employing over 20,000 workers in 1973 (Table 13.1) nine built their first large factory between 1902 and 1908.[5] Since then the Japanese company, Bridgestone, has been the only major new entrant into the global oligopoly.

In metals, group 34, the scale economies made possible by maintaining a high-volume throughput were also striking. Andrew Carnegie was able to reduce the cost of making steel rails by the new Bessemer steel process from close to $100 a ton in the early 1870s to $12 by the late 1890s.[6] In the refining of nonferrous metals, the electrolytic refining process patented in 1886 brought even more impressive cost reductions, permitting the price of a kilo of aluminum to fall from 87.50 francs in 1888 to 47.50 francs in 1889. After the new plants went into full operation, the price fell to 19 francs at the end of 1890, and 3.75 francs in 1895.[7]

In the machinery-making industries, groups 35–37, new technologies based on the fabricating and assembling of interchangeable metal parts were perfected in the 1880s. By 1886, for example, Singer Sewing Machine had two plants—one in New Jersey and the other in Glasgow, each producing 8,000 machines a week.[8] Maintaining an output that satisfied three-fourths of the world demand required an even more tightly scheduled coordination of flow of materials into, through, and out of the plant than did the mass production of packaged goods, chemicals, and metals. By the 1890s, a tiny number of enterprises using comparable plants supplied the world demand for typewriters, cash registers, adding machines, and other office equipment; for harvesters, reapers, and other agricultural machinery; and for the newly invented electrical and other volume-produced industrial machinery. The culmination of these processes came with the mass production of the automobile. By installing the moving assembly line in his Highland Park plant in 1913, Henry Ford reduced the labor time used in making a Model T from twelve and one-half man hours to one hour and thirty-three minutes.[9] This dramatic increase in throughput permitted Ford to drop the price of the touring car from over $600 in 1913 to $490 in 1914, to $290 in the 1920s; to pay the highest wages; to dominate the global market for low-priced automobiles; and to acquire one of the world's largest fortunes in an astonishingly short time.

On the other hand, in the SIC classification of Tables 13.1 to 13.5, the areas where few large firms appeared—that is, in the older, technologically simple, labor-intensive industries such as apparel, textiles, leather, lumber, and publishing and printing—neither technological nor organizational innovation substantially increased minimum efficient scale. In these industries, large plants did not offer significant cost advantages over small ones. In these industries, the opportunities for cost reduction through more efficient coordination of high-volume throughput by managerial teams remained limited.

The differences in potential scale economies of different production technologies indicate not only why the large hierarchical firms appeared in some industries and not in others, but also why they appeared suddenly in the last decades of the nineteenth century. Only with the completion of the modern transportation and communication networks—the railroad, telegraph, steamship, and cable—could materials flow into a factory or processing plant and the finished goods move out at a rate of speed and volume required to achieve substantial economies of scale. Transportation that depended on the power of animals, wind, and current was too slow, too irregular, and too uncertain to maintain a level of throughput necessary to achieve modern economies of scale.

However, such scale economies measured by throughput do not in themselves explain why the new technologies made possible by the new transportation and communication systems caused the new mass producers to integrate forward into mass distribution. Coordination might have been achieved through contractual agreement with intermediaries, both buyers and sellers. Such an explanation requires a more precise understanding of the process of volume distribution; particularly why the wholesaler, retailer, or other commercial intermediaries lost their cost advantage vis-à-vis the volume producer.

The intermediaries' cost advantage lay in exploiting both the economies of scale and what has been termed the "economies of scope." Because they handled the production of many manufacturers, they achieved a greater volume and lower per-unit cost than any one manufacturer in the marketing and distribution of a single product. Moreover, they increased this advantage by the broader scope of their operation; that is, by handling a number of related products through a single set of facilities. This was true of the new volume wholesalers in apparel, dry goods, groceries, hardware, and the like and even more true of the new mass retailers—the department store, the mail order house, and the chain or multiple shop enterprise.

The commercial intermediaries lost their cost advantages when manufacturers' output reached a comparable scale. As one economist pointed out: "The intermediary will have a cost advantage over its customers and suppliers only as long as the volume of transactions in which he engages comes closer to that [minimum efficient] scale than do the transactions volumes of his customers or suppliers."[10] This rarely happened in retailing, except in heavily concentrated urban markets, but it often occurred in wholesaling. In addition, the advantages of scope were sharply reduced when marketing and distribution required specialized, costly, specific facilities and skills that could not be used to handle other product lines. By investing in such product-specific personnel and facilities, the intermediary not only lost the advantages of scope but also became dependent on what were usually a small number of producers to provide those suppliers.

All the new volume-producing enterprises created their own sales organization to advertise and market their products nationally and in most cases internationally. From the start, they preferred to have a sales force of their own to advertise and market goods. Salesmen of wholesalers and other intermediaries who sold the products of many manufacturers, including those of their competitors, could not be relied upon to concentrate on the single product of a single manufacturer with the intensity needed to attain and maintain market share necessary to keep throughput at minimum efficient scale.

As important, mass distribution of these products often required extensive investment in specialized, product-specific facilities and personnel. This was because nearly all of the products were either entirely new, or greatly improved versions of existing ones. They were, to use an economist's term, highly differentiated. Some required costly, specialized distribution facilities if they were to be marketed in the volume at which they were produced. Others called for demonstration, installation, after-sales service and repair, extensive consumer credit, and other specialized marketing services. Wholesalers hesitated to risk the investment needed to handle such new and often untested products. Even when they were willing to accept that risk, they hesitated to make the costly investments in facilities and personnel that could be used for only a handful of specialized products processed by a handful of producers on whom they would become dependent for the supplies essential to make the investment pay. After all, their business and profits rested on handling related products of many manufacturers.

Of all the new mass producers, those making packaged food products and consumer chemical products—cigarettes, matches, breakfast cereals, canned soups, milk, meat, soap, pills, and paints—required the least in the way of product-specific distribution facilities and personnel. However, the new canning and packaging techniques did immediately eliminate one of the major functions of the wholesaler—that of converting large bulk shipments into small packages. Because the manufacturers now packaged, they, not the wholesaler, began to brand and then to advertise on a national and global scale. Their sales forces now canvassed the retailers. But because mass sales of these branded packaged products demanded little in the way of specialized facilities and personnel, their processors continued to use the wholesaler to distribute physically the goods for a fixed markup or commission until the manufacturers' output became large enough to cancel out the scale advantages of the wholesalers who handled the products of many firms.

All other industrial groupings in which the large firm clustered required major investments in either specialized distribution facilities or specialized personnel and often of both. The distribution of perishable products and those produced by the new mineral and vegetable oil refining processes required a massive investment in facilities if their sales were to be high enough to keep the processing plants at minimum efficient scale. In the 1880s, the meat packers built and operated fleets of refrigerated cars and established refrigerated storage facilities in every city and major town in the United States. Soon they had ocean-going refrigerated ships and storage areas in the ports of Britain and continental Europe.[11] Also in the 1880s, Standard invested in a fleet of railroad tank cars and storage depots as it built its national marketing organization. As its marketing became global, so too did its ocean-going tanker fleet and its intercontinental network of storage depots.[12] American Cotton Oil made comparable investments at home and abroad, including a major storage unit in Rotterdam to facilitate its European deliveries.

When the coming of the automobile required still another costly distribution investment in pumps and service stations to provide roadside supplies to motorists, wholesalers were even less enthusiastic about making the necessary investment. On the other hand, the refiners, by making the investment, were not only assured of being able to control the scheduling of throughput necessary to maintain their high minimum efficient scale but were more certain that the quality of the product was not adulterated, as it might have been if it were packaged by independent wholesalers. In the case of gasoline, in order to avoid the costs of operating the pumps and service stations, most oil companies preferred to lease the equipment they purchased or produced to franchised dealers. A comparable situation occurred in tires, whose mass production benefited from the economies of throughput and whose mass sales required a specialized product-specific distribution network. Again, although tire companies occasionally owned their retail outlets, they preferred to rely on franchised retail dealers.

The mass marketing of newly invented machines that were mass-produced through the fabricating and assembling of interchangeable parts required a greater investment in personnel to provide demonstration, after-sales service and repair, and consumer credit than that required for product-specific facilities.[13] The mass distribution of sewing machines (for households and for the production of apparel), typewriters, cash registers, adding machines, mimeograph machines, and other office equipment; harvesters, reapers, and other agricultural machines, and, after 1900, automobiles and the more complex electrical appliances all required such services. Because these machines were so new, few existing distributors had the necessary training and experience to provide these services or the financial resources to extend long-term consumer credit.

On the other hand, the manufacturer had every incentive to do both. The provision of repair and service helped assure the product performed as advertised. As important, if they owned and operated a wholesale organization they were more certain of assured inventory as well as quality control. Moreover, as many retailers were needed to cover the national and international markets, these machinery manufacturers preferred to rely on, as did the oil and tire companies, franchised dealers. The retail dealers, who sold their products exclusively, were supported by a branch office network that provided the credit, delivered supplies on schedule, and had trained mechanics and special equipment to assure after-sales service and maintenance. Only the makers of sewing machines, typewriters, and cash registers went so far as to invest in retail stores. They did so primarily in concentrated urban areas where, before the coming of the automobile, only such stores were able to provide the necessary services and credit on a neighborhood basis.

The makers of heavier but still standardized machinery for industrial users often required even more in the way of demonstration, installation of the new machines, continuing maintenance and service, and consumer credit. In providing such service to purchasers of these technologically complex new or greatly improved shoe machines, turbines, pumps, boilers, elevators, telephone equipment, and the machinery to generate, transmit, and use electric power, manufacturers' agents and other intermediaries rarely had the training or the capital required. For the makers of industrial chemicals, volume distribution demanded investment in product-specific distribution facilities, as well as salesmen with specialized skills. Dynamite, far more powerful than black powder, required careful education of customers, as well as specialized storage and transportation facilities. So, too, did the new synthetic dyes and synthetic fibers whose use had to be explained to manufacturers and whose application often required new specialized machinery. On the other hand, metals produced by processes with a high minimum efficient scale required less investment in distribution. Even so, to obtain and fill volume orders to precise specifications on precise delivery schedules called for a trained sales force and for close coordination between production and sales managers.

It was this investment in costly, product-specific distribution facilities that carried the pioneering volume-producing manufacturers of new or greatly improved products abroad, that is, that led the large industrial firm to become multinational by making direct investments abroad in marketing. For a U.S. manufacturer located, let us say, in New York, little more effort was needed to set up branch offices with

distribution facilities and personnel in Britain and Europe than it was to do so on the U.S. West Coast. Where transportation or tariffs substantially increased cost and where local markets were large enough to justify the building of plants sizable enough to achieve the cost advantages of scale, these companies built, or occasionally purchased plants to supply their marketing organizations. If these foreign markets continued to expand, the distant subsidiary often began to purchase its supplies as well as selling finished products locally. In this way, the foreign subsidiary became as integrated as its parent.

In nearly all the new volume-producing industries, the pioneers continued to dominate for decades. Investment in plants of minimum efficient scale, the building of a marketing and distribution network, and the recruitment of managerial hierarchies to administer and coordinate production and distribution created powerful barriers to entry. Late comers had to build plants of comparable size if they were to be competitive in costs and to set up comparable sales forces if they were to have a share of the market large enough to maintain that scale. Moreover, where the first movers were often able to finance continued expansion from retained earnings, the newcomers had to finance a much larger initial investment in both production and distribution and therefore paid higher financing costs. Finally, the new set of managers had to learn their trade while competing with experienced rivals. Competitors did appear, but they were few in number. From almost their beginnings these industries were oligopolistic. In most cases the oligopoly was international.

In these ways and for these reasons, the large industrial firm that integrated mass production and mass distribution appeared in industries with two characteristics. The first and most essential was a technology of production in which the realization of potential scale economies and maintenance of quality control demanded close and constant coordination and supervision of material flows by trained managerial teams. The second was that volume marketing and distribution of their products required investment in specialized, product-specific human and physical capital.

Where this was not the case, that is, in industries where production technology did not have a potentially high minimum efficient scale, where coordination was not technically complex, and where mass distribution did not require specialized skills and facilities, there was much less incentive for the manufacturer to integrate forward into distribution. In such industries as publishing and printing, lumber, furniture, leather, and apparel and textiles, the large integrated firm had few competitive advantages. In these industries, the small, single-function firm continued to prosper and to compete vigorously.

In industries that had the two critical characteristics, the most important entrepreneurial act after the initial investment in the new production technology was the creation of an administrative organization. This entailed the recruitment of a team to supervise the process of production, then the building of a national and normally international sales network, and finally the setting up of a corporate office of middle and top managers to integrate and coordinate the two. Only then did the enterprise begin to make direct investment abroad in production facilities. Such investment followed—almost never preceded—the building of an overseas marketing network. So, too, in the technologically advanced industries, the investment in research and development came after, not before, the creation of a marketing network. In these firms, the linkage between trained sales engineers, production engineers, product designers, and the research laboratory became a major impetus to continuing innovation in the industries in which they operated. The result of such growth was an enterprise whose organization is depicted in Figure 13.1. The continuing growth of the firm rested on the ability of its managers to transfer resources in marketing, research and development, and production (usually those that were not fully utilized) into new and more profitable related product lines, a move that carried the organization shown in Figure 13.1 to that illustrated by Figure 13.2. If the first step—that of integrating production and distribution—was not taken, the rest did not follow. The firms remained small, personally managed producing enterprises buying their materials and selling their products through intermediaries.

National Differences in the Growth of the Multinational Enterprise

The large managerial enterprise evolved in much the same way in industries with much the same characteristics in the major modern economies. However, there were striking differences in the pace, timing, and specific industries in which the new institution appeared and continued to grow in each of these economies. These differences reflected differences in technologies and markets available to the industrialists of the different nations, differences in their entrepreneurial and organizational skills, and differences in laws, cultural attitudes, and values. These dissimilitudes can be pinpointed by very briefly reviewing the historical experiences of the two hundred largest industrial enterprises in the United States, the United Kingdom, Germany, and Japan.[14]

The United States

In the United States, the completion of the nation's basic railroad and telegraph network and the perfection of its operating methods in the 1870s and 1880s opened up the largest and fastest growing market in the world. Its population, which already enjoyed the highest per capita income in the world, was equal to that of Britain in 1850, twice

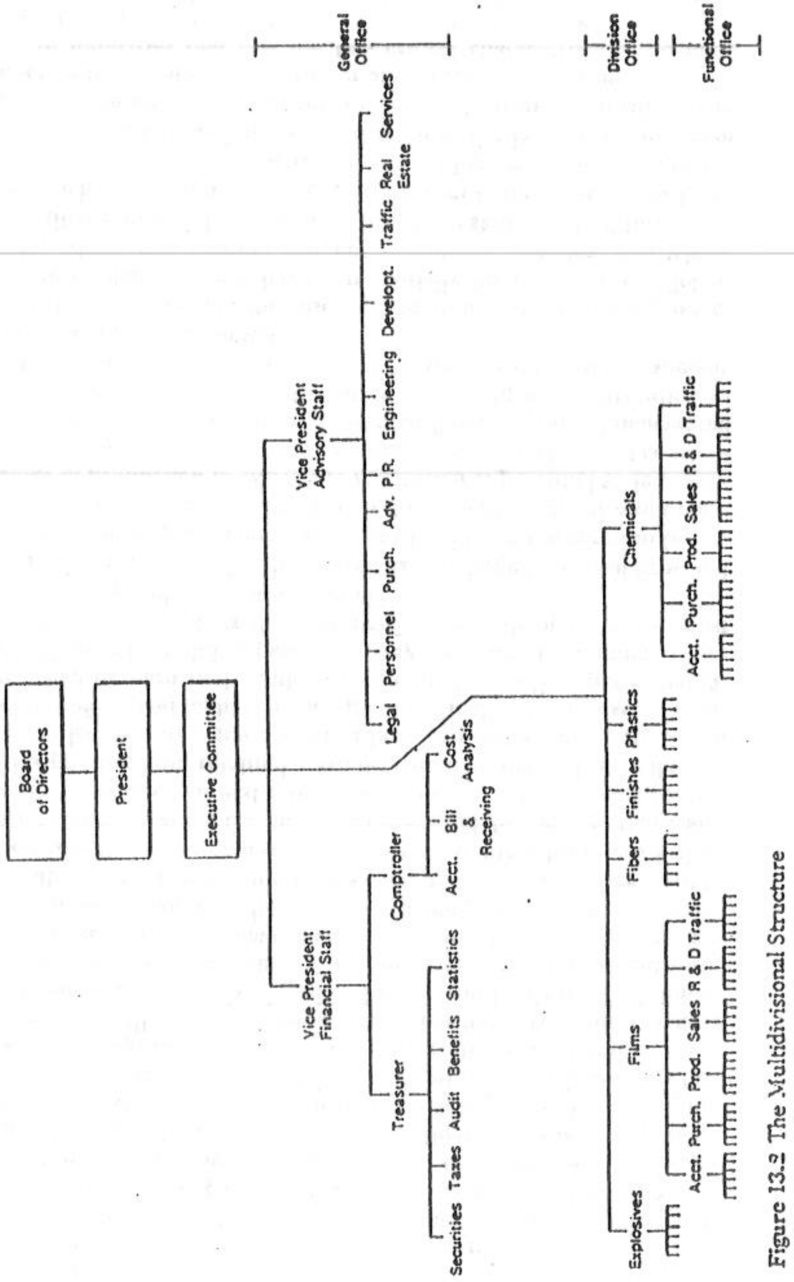

Board of Directors
President
Executive Committee
Vice President Advisory Staff
Legal
Personnel
Purch.
Adv.
P.R.
Engineering
Developt.
Traffic
Real Estate
Services
Vice President Financial Staff
Treasurer
Securities
Taxes
Audit
Benefits
Statistics
Comptroller
Acct.
Bill & Receiving
Cost Analysis
Explosives
Films
Fibers
Finishes
Plastics
Chemicals
Acct.
Purch.
Prod.
Sales
R & D
Traffic
General Office
Division Office
Functional Office

Figure 13.2 The Multidivisional Structure

Table 13.6. Multinational Companies in 1914[a]

Groups 20 & 21: Food and Tobacco

American Chicle
American Cotton Oil
Armour
Coca-Cola
H. J. Heinz
Quaker Oats
Swift
American Tobacco
British American Tobacco

Groups 28, 29, & 30: Chemicals & Pharmaceuticals, Oil and Rubber

Carborundum
Parke Davis (drug)
Sherwin-Williams
Sterns & Co. (drug)
United Drug (drug)
Virginia-Carolina Chemical
Du Pont
Standard Oil of N.J.
U.S. Rubber

Groups 35, 36, & 37: Machinery and Transportation Equipment

American Bicycle
American Gramophone
American Radiator
Chicago Pneumatic Tool
Crown Cork & Seal
Fort Motor
General Electric
International Harvester
International Steam Pump (Worthington)
Mergenthaler Linotype
National Cash Register
Norton
Otis Elevator
Singer Sewing Machine
Torrington
United Shoe Machinery
Western Electric
Westinghouse Air Brake
Westinghouse Electric

Others

Alcoa (33)
Gillette (34)
Eastman Kodak (38)
Diamond Match (39)

[a]U.S. companies with two or more plants abroad or one plant and raw material producing facilities.
SOURCE: Mira Wilkins, *The Emergence of Multinational Enterprise* (Cambridge, Mass.: Harvard University Press, 1970) 212–13, 216.

that in 1900, and three times that in 1920.[18] U.S. entrepreneurs quickly made the investment in plants large enough to exploit the cost advantages of scale, recruited the managerial teams in production necessary to exploit scale economies, and made the investment in distribution necessary to market their volume-produced goods at home and abroad. They did so in all the industries in which large industrial firms would cluster for the following century. Most of these firms quickly extended their marketing organizations overseas, and then became multinational by investing in production facilities abroad, playing an influential role in a global oligopoly. Table 13.6 lists the U.S. industrial firms with production facilities abroad in 1914. By that date, the value of direct foreign investment by U.S. multinationals equaled 7 percent of the gross national product. That figure was still the same in 1966.[19] In some cases, particularly in mass-produced light machinery, the United States enjoyed close to global monopoly well before the outbreak of World War I. By that time too, multinationals in the more technologically advanced industries had begun to invest personnel and facilities in R & D.

In industries where several entrepreneurs simultaneously adopted new technologies to produce new or greatly improved products, the large enterprise initially came into being through mergers and acquisition. In such industries, manufacturers both in Europe and the United States responded to excess production capacity by cooperating with each other to control competition through a variety of contractual forms. U.S. manufacturers first attempted such cooperation by forming trade associations to control output and prices or to allocate marketing territories.[17] However, because of the existing common law prohibition against combinations in restraint of trade, these associations were unable to enforce their rulings in courts of law. So manufacturers turned to the holding company device. Members of their association exchanged their stock for that of a holding company, thus giving the central office of the new holding company legal power to determine output, prices, and marketing areas for the subsidiary firms.

However, for some, like John D. Rockefeller, such legal consolidation became the first step for rationalizing the resources of an enterprise or even an industry in order to exploit more fully potential scale economies. Even before the enforcement of the Sherman Antitrust Law in the early twentieth century made contractual cooperation by means of a holding company legally suspect, a number of U.S. enterprises had been transformed from a holding company into an operating one, by consolidating the factories of their subsidiaries into a single production department and unifying the several sales forces into a single sales department. Rationalization of personnel and facilities followed legal consolidation and administrative centralization. Some plants were shut down, others enlarged, and new ones were built in a way to exploit most effectively the cost advantages of scale in production. The sales network was reshaped to handle existing and forecasted sales requirements. Then, though less often than not, investment was made in research and development. In this way, a loose federation of operating companies was transformed into a single centralized enterprise as depicted on Figure 13.1. Whether an enterprise reached this form after merger and acquisition or by continuing reinvestment in its own production, distribution, and research facilities and personnel, these large managerial enterprises had become firmly entrenched in major sectors of the U.S. economy by the coming of World War I.

Great Britain

This was certainly not the case in Great Britain. As late as World War I, the large integrated industrial enterprise administered through an extensive managerial hierarchy was still the exception. Nearly all of the two hundred leading industrials in Britain had integrated production with distribution, but in a great number of these, owners remained full-time executives. They managed their enterprises with the assistance of a small number of "company servants" who only began to be asked to join boards of directors in the 1930s. In Britain, at the time of World War II, most of the top two hundred consisted of two types of enterprise, neither of which existed among the U.S. top two hundred at the time of World War I. They were either personally managed enterprises or federations of such enterprises. The exceptions were, of course, Britain's largest and best-known industrial corporations—those that represented Britain in their global oligopolies. However, as late as 1948, these numbered less than 20 percent of the top two hundred enterprises.

Large hierarchical enterprises did come when British entrepreneurs responded to the potential of new high-volume technologies by making the necessary investment in production facilities, creating management teams for production, and investing in distribution and research personnel and facilities. Between the 1880s and World War I, such individual entrepreneurs responded in this way in the production of several branded packaged products (like soap, starch, biscuits, and chocolates) as well as rayon, tires, plate and flat glass, explosives, and synthetic alkalis. For example, Courtaulds, the first to build a plant with a high minimum efficient scale in rayon, became and remained the largest producer of the first synthetic fiber, not only in Britain but also in the United States.

But where British industrialists failed to grasp the opportunity to make the investment and build the hierarchies, they not only lost the world market but the British home market itself. This was particularly striking in machinery, both light and heavy, and in industrial chemicals. The U.S. firms quickly overpowered the British competitors in the production and distribution of light mass-produced machinery, including sewing, office, and agricultural machinery, automobiles, household appliances, and the like. The Germans as quickly dominated the synthetic dye business so critical to Britain's huge textile industry, while both the Germans and Americans moved into Britain's electrical machinery industry. In 1912, two-thirds of the output of the electrical equipment industry within Britain was produced by the British factories of the U.S. General Electric and Westinghouse and the German Siemens.[18] Even those few firms that achieved and maintained their position in the British market and the global oligopoly created smaller hierarchies and had more direct owner management than did their U.S. counterparts.

After World War I, a few British firms in these new capital-intensive industries began to challenge their U.S. and German competitors, but they did so only by making the necessary investment in nonmanufacturing personnel and facilities and by recruiting managerial staffs. This was the case for Anglo-Persian Oil Company (it became British Petroleum), for British General Electric, and Imperial Chemical Industries in each of their industries, for Metal Box in cans and for Austin and Morris in automobiles. Nevertheless, the transformation from personal or family management to one of salaried managers came slowly and grudgingly. Even in the largest enterprises—those with sizable hierarchies such as Courtaulds, British Celenese, Pilkington, Metal Box, Reckitt, Cadbury, Rank, and others—the owners continued to have a much greater say in the top management decisions than did their U.S. counterparts.

Why was this the case? The answer is complex. It lies in Britain's industrial geography and history, in its educational system, in the lack of antitrust legislation, and in a continuing commitment to personal and family management. Because Britain's domestic market was smaller and growing more slowly than that of the United States, British industrialists had less incentive than their U.S. counterparts to exploit scale economies. Moreover, Britain was the only nation to industrialize before the coming of modern transportation and communication. So its industrialists had become tuned to a slower, smaller-scale process of industrial production and distribution.

Nevertheless, precisely because it was the first industrial nation, Great Britain also became the world's first consumer society. The quadrangle bounded by London, Cardiff, Glasgow, and Edinburgh remained for almost a century after 1850 the richest and most concentrated consumer market in the world. British entrepreneurs quickly began to mass produce branded packaged consumer goods. (Of all the new industries these required the least in the way of specialized skills in production and specialized services and facilities in distribution.) But in other of the new industries, it was foreign, not British, entrepreneurs who responded to the new opportunities. Even though that golden quadrangle remained the world's most concentrated market for mass-produced sewing machines, shoe machinery, office equipment, phonographs, batteries, automobiles, appliances, and other consumer durables as well as electrical and other new heavy machinery and industrial chemicals, Germans and Americans were the first to make *within Britain* the large investment required to exploit the new technologies fully, to establish the product-specific distribution services and facilities, and to recruit the essential managerial staff. Apparently British industrialists wanted to manage their own enterprises rather than turn over operating control to nonfamily, salaried managers. They seemed to regard their companies as family estates to be nurtured and passed down to their heirs rather than mere money-making machines. As a result, they and the British economy as a whole failed to harvest many of the fruits of the second industrial revolution.

The commitment to family control was reflected in the nature of British mergers. As in the United States, many British firms grew large by merger and acquisition. As in the United States, holding companies were formed to control legally the output, price, and marketing arrangements of hitherto small competing enterprises; but in Britain, unlike in the United States, holding companies remained federations of family firms. Until after World War I, no British industrialist viewed merger as forerunner to the rationalization, consolidation, and centralized administration necessary to exploit the potential scale of economies. Indeed, the very first merger to centralize and rationalize in Britain came in 1920 at Nobel Explosives, the forerunner to ICI, which borrowed the necessary organizational techniques directly from its overseas ally, the Du Pont company of Wilmington, Delaware.[19] As late as 1928, Lever Brothers, one of Britain's largest enterprises, had forty-one operating subsidiaries and thirty-nine different sales forces. For these reasons, then, the founders of most large British enterprises continued to manage their enterprises directly. Hierarchies remained small and controllable. Sons, grandsons, and grandsons-in-law continued to move into the top offices. Britain continued until World War II to be the bastion of family capitalism.

Germany

In Germany, unlike Britain, there were integrated industrial firms that were as large as those in the United States well before the coming of World War I. They were, however, fewer in number and were concentrated in metals and the technologically advanced machinery and chemical industries. Among the top two hundred German firms during the interwar years, there were very few producing branded packaged products, except for the regional breweries. One can only locate two chocolate and confectionary and two drug companies among the top two hundred. The remaining few were subsidiaries of Nestlé's, Lever Brothers, and the two Dutch margarine makers that joined Lever in 1929 to become Unilever. Nor did the large German firms manufacture light mass-produced machinery in the U.S. manner. Singer Sewing Machine long remained the largest sewing machine maker in Germany. Well before World War I, the factories of National Cash Register and American Radiator and the sales office of International Harvester and Remington Typewriter dominated the German market for their products. In automobiles in 1929, the year when General Motors produced 1.6 million and Ford 1.5 million cars, only one Germany car company made more than 10,000, and that firm, Adam Opel, which produced 25,000, had just become a General Motors subsidiary. Even in standardized industrial machines, U.S. firms such as Mergenthaler Linotype (in printing presses) and Norton and Carborundum (in abrasives and grinding machines) dominated German markets.

The Germans did, like the British, have their one large representative in the global rayon, rubber, and oil oligopolies. (The last, EPU, was dismembered during World War I.) It was in complex machinery and chemicals, however, that the Germans made their global mark. In giant production works, German machinery and chemical enterprises produced in quantity a variety of complex machines and chemicals. Managerial hierarchies even greater in size than those of the production departments of U.S. firms guided the complicated flow of materials from one intermediate process to the next. In the 1880s and 1890s, enterprises—such as Bayer, Hoechst, BASF, AGFA, and Degussa in chemicals; Merck, von Heyden, and Schering in pharmaceuticals; Siemens and AEC in electrical equipment; AFA in storage batteries; and Metallgesellschaft in copper—built extended networks of branch offices throughout the world to market their products, demonstrate their use, install them where necessary, provide continuing after-sale service, and give customers the financial credit they often needed to make such purchases. Once established abroad, they built or acquired branch factories. Finally, they invested, usually more heavily than the Americans in research and development.

At home these large integrated enterprises reduced competition by making contractual arrangements for setting price and output and allocating markets. Because such arrangements were in Germany legally enforceable in courts of law, the arrangements became quite formal and elaborate. The I.G., a legal form that is best translated as the "community of interest," became the closest to that of the British and the U.S. holding company. The difference between the British holding company and the German community of interest was that the latter involved large hierarchical firms rather than small family enterprises. Their extensive investment in marketing and distribution and in research and development permitted these firms to dominate the negotiations setting up cartels or associations both at home and abroad and provided them with the power essential to implement and enforce the contractual arrangements.

Finally, the capital requirements of these capital-intensive producers of industrial products were far greater than those of the U.S. and British makers of branded packaged products or the U.S. mass producers of light machinery. Because there were no highly developed capital markets in Germany comparable to those of London and New York, German banks became much more involved in the financing of large hierarchical enterprises than was true in Britain and the United States. Although the representatives of banks rarely sat on the *Vorstand*, the central administrative body of top managers, as did the founder and often full-time family executives, they did become important members of the *Aufsichtsrat*, or supervisory board.

Why was it that the large German industrial enterprises were concentrated in metals and complex industrial products rather than branded goods or light mass-produced machinery? Why did the Germans build large hierarchical organizations when the British did not?

Manufacturers in the new German empire of the 1870s, at the time when the transportation and communication revolution was nearing an end, enjoyed neither the rapidly growing continental market of the United States nor the concentrated consumer market of Britain. Because per capita income was lower in Germany than in the two Anglo-Saxon countries, and because Germany was neither a large importer of foodstuffs like the United Kingdom nor an exporter like the United States, neither the markets nor the availability of supplies for mass producing and mass distributing branded packaged products was comparable to those in Britain and the United States. On the other hand, the rapidly growing industrial markets in Europe, including Britain and Germany itself, created a powerful growing demand for the new specialized industrial machinery, including electrical equipment, and new industrial chemicals, including synthetic dyes. In building their technical sales and research organizations—their basic weapons in international competition—the Germans had the advantage of what had become the world's best technical and scientific educational institutions. Therefore, despite the defeat in two wars, the German strength in international competition still rests on the performance of their science-based industries.

Japan

The evolution of the large industrial enterprises in Japan differed greatly from those in the West. For Japan was just taking the first steps toward modern industrialization in the same decades that the new transportation and communication revolution was spawning the second industrial revolution in Europe and the United States. Indeed, Japan's first steel mill only went into operation in 1902. Only in the years after World War II was the economy large and strong enough to support modern mass production and mass distribution. Yet even before that war managerial hierarchies had appeared to exploit new technologies and to reach new markets.

In the early years of this century, Japan's domestic and foreign markets were totally different. At the time of the Meiji Restoration, Japanese manufacturers enjoyed a highly concentrated domestic market, comparable to Britain's during its early industrialization, with long-established channels for distribution of traditional consumer goods. As a result, only a few Japanese firms (and no foreign companies) began to create marketing networks to distribute branded packaged products within the country. By World War II, a small number of makers of branded packaged products such as confectionary, soy sauce, canned sea food, beer, and soap, who advertised nationally and had their own extensive sales forces, were listed among the largest two hundred Japanese industrial enterprises.

On the other hand, the Japanese had had no overseas commercial contact at all, even with the nearby Asian continent, for the more than two hundred fifty years of the Tokuagawa rule. Manufacturers using imported processes to produce textiles, fertilizers, ceramic, and metal products sought overseas as well as domestic markets, particularly in nonindustrialized East and Southeast Asia. Overseas they rarely set up their own branch offices. They had neither the volume nor the distribution needs to require large product-specific investments in distribution. They relied instead on allied trading companies to assure coordination of the flow of goods from factories to Japan to customers abroad, and also to those at home, and the flow of essential materials and equipment from overseas to the producing facilities. These trading companies set up branch offices in Japan and in all parts of the world as well as building a large central office in Tokyo or Osaka. That is, they invested in an extensive marketing and distribution organization that coordinated flows, provided marketing services, and generated information, thus lowering marketing and distribution costs. The trading companies became the linchpins of groups of single-product manufacturing enterprises, each group having its own banks and trust companies, as well as its own trading and warehouse concerns.

The close relationship between the managers of the manufacturing companies and those of the trading firms, either within the giant *zaibatsu* or between cooperating manufacturers in less formal groups, permitted the Japanese to capture an increased share of world trade, particularly in the relatively low-technology industries. However, where marketing and distribution did require product-specific skills, services, and facilities, enterprises set up their own distributing network and operated outside of the *zaibatsu* and other group enterprises. Before World War II, only a small number of such enterprises had appeared, primarily in industrial machinery, particularly electrical. The latter was especially important, for until the 1950s, Japan relied heavily on hydroelectric power for its energy. Only after the war, with the rapid growth of the domestic market did the makers of automobiles, electric appliances, radio, and television build comparable organizations. In the postwar years, these enterprises increasingly invested in distribution abroad and operated through extensive managerial hierarchies comparable to those of the West.

GLOBAL COMPETITION DURING THE INTERWAR YEARS

This review of the growth of the modern integrated industrial multinational provides the data necessary to understand the changing nature of competition among such enterprises in international trade. These changing patterns of competition can be most precisely presented by reviewing, first, competition during the years between World War I and World War II and, then, by examining the ways in which changing technologies and markets within a changing system of nation states helped to transform such competition in the post–World War II years.

In the international markets of the interwar years, even after the onslaught of the Great Depression, multinationals played an increasingly significant role. They grew both in numbers and in size. In these years U.S. multinationals continued to be more numerous in more industries and in some industries larger in size than those of other nations. They divided fairly evenly between those producing and distributing consumer goods and those handling industrial products. British multinationals concentrated more on the production of consumer goods, particularly branded packaged products; while the strength of the German multinationals lay in the technologically advanced industrial chemicals, machinery, and metal products. The Japanese still played a minor role in international competition.

In these industries the first movers continued to maintain their competitive advantage. But challenges did appear, and market share in these global oligopolies rarely remained stable. The Americans continued to dominate in light machinery, becoming even stronger in appliances and automobiles. They remained strong in heavy machinery and also in tires.[20] In oil the successors to Standard Oil (which was dismembered in 1911 by a Supreme Court antitrust decision) were increasingly challenged by Anglo-Iranian Oil (British Petroleum), Royal Dutch Shell, Texaco, and Gulf. In branded packaged products, such companies as Corn Products, General Foods, National Biscuit, Wrigley Chewing Gum, Procter & Gamble, Colgate, and American Home Products made direct investments in manufacturing to support their sales organizations. The major expansion of British multinationals came in the production of branded packaged products including consumer chemicals.[21] They continued to concentrate on the European and imperial markets. The Germans were severely hurt by World War I, for the British, French, and Americans expropriated subsidiaries of their chemical, pharmaceutical, electrical and nonelectrical machinery, copper, and oil multinationals. Nevertheless, most of these firms quickly recovered from wartime losses. They soon began to regain market share, particularly in Eastern and Southern Europe but also in the wider international markets including chemicals in the United States.[22] Nevertheless, the confiscation of overseas manufacturing and other properties during World War I discouraged them from making capital investments abroad in the interwar years comparable to those made before 1914. In Japan, whose trade had benefited greatly from World War I, the trading companies of the *zaibatsu* not only expanded their activities primarily in the nearby markets of East Asia, but also enlarged the size and number of their branch offices in the United States and Europe.[23]

The growing economic autocracy of the leading nation states during the interwar years encouraged expansion abroad. Tariff barriers became increasingly higher and trade restrictions and currency controls increasingly rigid. Nevertheless, the desire to protect existing markets and to move in the new ones probably remained an even more powerful motive than tariff jumping for direct investment abroad.

In these years, differences in the ways of international competition reflected not only those in the technologies of production, but also differences in the patterns of competition and cooperation developed in various domestic markets. These differences in competition at home in turn reflected differences in the size of the domestic markets and in national laws affecting interfirm competition. In the United States, the large and rapidly growing domestic market permitted several enterprises to operate plants at minimum efficient scale. In Europe, on the other hand, domestic markets could rarely support more than one or two such enterprises. U.S. government legislation penalized monopoly but not oligopoly. European governments remained much more lenient toward monopoly.

Oligopolistic competition in the United States was rarely based on price alone. Usually the largest, which was often the oldest, became the price leader. The firms did, however, compete, often forcefully, for market share and profits by what can be termed functional and strategic efficiency. That is, they competed by becoming more efficient in carrying out the several processes of production and distribution, by developing and improving both product and process through systematic research and development, by locating new and more suitable sources of supply, and by providing more effective marketing services and carrying out more forceful marketing campaigns. With such functional efficiencies they maintained or improved their competitive advantages of cost and product differentiation. In addition, by more careful planning and resource allocation they moved more quickly than competitors into new and expanding markets and out of old and declining ones.[24]

Antitrust legislation encouraged this form of competition. As interpreted and enforced by the Federal courts, monopoly remained illegal; but oligopoly was sanctioned, depending in part on what the Supreme Court determined to be reasonable competition. The new competition did result in gains and losses, occasionally quite dramatic ones, in market share and in rate of return for the leaders in such major industries as steel, oil, rubber, glass, agricultural machinery, meat packing, and automobiles during the years between the turn of the century merger movement and World War II.

In Europe, much smaller domestic markets and the lack of antitrust legislation meant that market power was achieved and maintained in the domestic market far more by contractual cooperation than through functional and strategic efficiencies. In those British In-

873
625

dustries where a single firm did not dominate, federations of relatively small, usually family enterprises, normally in the form of holding companies, maintained agreements as to price, output, and marketing territory. In Germany, where cartel agreements were legal and could be enforced in courts of law, market control through contractual cooperation was even more effective.

In international competition, however, such contractural cooperation was much more difficult to enforce. No international courts existed to bring cheaters on the cartel into line. Nor was the holding company as effective across borders as it was within a single national legal system. Thus, by the interwar years, international competition became patterned more on the U.S. style of oligopolistic competition than on the European brand of monopoly. Nevertheless, in industries, particularly those producing industrial goods where technologies of production brought an extremely high throughput and so permitted a very small number of firms to dominate world production and where these few were primarily European firms, global cartelization did occur, but rarely on a permanent basis.

Industry Differences in Interwar Global Competition

During the interwar years, the ways of reflected differences in technologies of production in different industries as well as differences in national laws and traditions relating to competition helped to shape the organizational structure of the competing firms. In branded packaged food, tobacco, and chemical products (groups 20 and 21 and consumer products in group 28) the United States and Britain continued to dominate international markets. The Americans appear to have been more agressive in competing in Britain and the Continent than were the British in the United States. The U.S. companies' abilities to increase market share abroad rested on the transfer of marketing and distribution skills developed at home. The British preferred to concentrate more by expanding within their empire, that is, India and particularly the white commonwealth countries where they sold to transplanted Britishers.[25] The British may have been handicapped in direct competition with the Americans because their firms lacked a strong central office to direct the development and transferal of specialized marketing and distribution skills.

In these consumer goods industries, scale economies were less pronounced. Therefore, there was less competitive advantage to concentrating production in large plants, and local tastes directly affected sales because there were relatively few major advantages in coordinating flows on a global scale. Both British and U.S. multinationals producing such branded packaged products operated through decentralized federated structures. The British firms that had smaller corporate offices gave less direction to their foreign subsidiaries and monitored their activities less carefully than did the Americans. Within their local, usually national, markets they competed functionally and strategically. Although some of the large British firms worked together closely in obtaining supplies, and even in sharing the domestic market, there was little evidence of contractural cooperation in the packaged goods industries among international rivals in the marketing and distribution of their products abroad.

In the oil and rubber industries (groups 29 and 30) scale efficiencies resulting from administrative coordination of the flow of relatively homogeneous products to and from refineries and processing plants brought central control. In oil the global oligopoly was quickly dominated by three firms. By the early 1920s, Anglo-Iranian (British Petroleum) and Shell created managerial hierarchies comparable to those of Standard Oil (New Jersey, now Exxon).[26] By the 1930s, four other U.S. firms were competing in the oligopoly. These global leaders had heavy investments in pipelines, tanker fleets, and railroad cars, the use of which was carefully coordinated. Because of their much greater scale economies, oil companies were slower to build overseas processing plants than were the producers of branded packaged products. When they did, the output of one foreign refinery tended to serve several nations in one large region. Once established, decisions of local managers had to be integrated with those of the corporate staff at the home office if production and distribution efficiencies were to be assured. The following statement on the functions of the corporate Coordination Department at Jersey Standard written in October 1927 makes this point effectively.

> In general, the Department assists in currently determining the policies to be adopted so far as the activities of the producing, transportation and manufacturing branches of the Company are concerned, and cooperates with the subsidiary companies to see that these policies are carried out. This entails arranging for refinery Crude supplies; Crude transportation and allocation; production, purchases and movements of certain finished products (i.e., naphthas [gasoline], refined oil, gas, and fuel oils); coordination of refining manufacturing operations from the standpoint of Crudes run and products required, stocks on hand, etc., and the allocation of business among the various subsidiary companies.[27]

After spelling out the complexities of global coordination in crude oil production, transportation, and refining, the memorandum reviewed the intricacies of coordinating globally the distribution of refined products.

Because of comparable scale economies in the production of tires and to a lesser extent in rubber boots and shoes, the pattern was somewhat the same in rubber. Members of the global oligopoly—four U.S. firms (Goodrich, Goodyear, Firestone, and U.S. Rubber) and one from each major industrial nation—concentrated production in a few plants, brought raw materials from vast distances and marketed worldwide. As the volume of output was smaller and sources of supplies and centers of distribution somewhat fewer, coordination in rubber required less careful scheduling than it did in oil. Nevertheless, the overall corporate structure was much more a centralized than federated one.

In both industries during most of the interwar years, these large mass producers of primarily consumer goods competed functionally and strategically for market share in the U.S. style. In the late 1920s, the opening of vast new oil fields in Texas, California, and Iran threatened overproduction of crude oil. The three largest companies, Jersey Standard, Shell, and Anglo-Iranian, signed the well-known Achnacarry, or "as is" agreement, which attempted to maintain stability by reducing competition for market share.[28] As the Great Depression dampened demand, the smaller of "the Seven Sisters"—Texaco, Gulf, Standard Oil of New York (now Mobil), and Standard Oil of California—joined the enlarged general agreement. Nevertheless, few specific arrangements as to actual price and output were reached and still fewer of those created enforcing agencies.

No such agreements were attempted by the rubber manufacturers. The only significant agreements in that industry were between rubber-producing nations that restricted crude rubber production. These agreements were strongly opposed by the members of the rubber oligopoly as it forced those that had not already integrated backward into owning and operating distant rubber plantations to do so.

In mass-produced light machinery, where the United States continued to dominate international markets until well after World War II, there were even fewer attempts at contractural cooperation. U.S. makers of sewing machines, office and agricultural machinery, automobiles, and such appliances as washing machines, vacuum cleaners, and electric stoves relied on their own extensive marketing organizations to demonstrate, install, service, repair, and provide consumer credit. By 1930, nearly all had enlarged the plants they had operated before the war. Many built and purchased new ones. When competition did come, as it did in automobiles in Britain, it came usually because of the failure of the U.S. home office to adjust to basic local market requirements.

In the fabricating and assembling industries (groups 35, 36, and 37) the choice of direct investment abroad in production often rested on whether scale economies were greater in the fabricating or in the assembling of the parts. In automobiles, and to some extent in agricultural and light construction equipment, scale economies were best achieved by concentrating fabricating in a small number of large plants, usually at home, while having the assembling of parts scattered in a large number of smaller plants closer to markets. To operate both fabricating and assembling plants at minimum efficient scale required careful centralized coordination of product flow. On the other hand, in the mass production of sewing machines, typewriters, and other office equipment, such economies in fabricating were less pronounced than in automobiles, so manufacturers normally built plants abroad that both fabricated and assembled. In such cases, production was usually concentrated in a very few large plants that sourced from and sold to large usually multinational regions. Therefore, central coordination was less essential. The regional subsidiary coordinated product flow with little consultation from the home office. However, basic product and process design as well as continuing research on process and product continued to be carried out in the home country, usually close to corporate headquarters.

Even where the coordinating role of the home office was more critical to achieving production economies in both fabricating and assembling, the differing strategies of the different firms affected their organization structure. Thus, coordination by the Detroit office at Ford, where growth resulted from direct investment in assembling plants, was tighter than it became at General Motors, which in the late 1920s shifted to an expansion strategy of purchasing existing integrated enterprises.[29] Nevertheless, General Motors's corporate officers in Detroit determined manufacturing, design, and product policies for its major foreign subsidiaries, Opel in Germany and Vauxhall in Britain.[30]

In the production and distribution of industrial goods—chemical, electrical, and other heavy machinery—where production normally required the coordination of several intermediate processes, the Germans and Americans who dominated the international markets usually unified these processes within a single large plant or works. Therefore, when they went abroad, they normally set up a complete works that sourced and sold regionally with managers of the subsidiary handling day-to-day activities. However, in these industries the corporate office continued to control basic price and output policies more tightly than those producing consumer products.

The volume producers of industrial goods attempted much more than did producers of automobiles, light machinery, rubber, and oil to control international markets through cartels. In electrical equipment the four leaders of the global oligopoly, two U.S. and two German firms, had patents and licensing agreements with each other and with smaller companies in other nations—agreements that allocated markets as well as pooling patents and providing for technical information. The same was true for German and U.S. producers of storage batteries and other more specialized electrical products. Nevertheless, because the market for such equipment was until the 1930s growing rapidly and because the opportunity for technological innovation remained high in both product and process, the smaller, although still quite large, British, Swiss, and Swedish electrical manufacturers were able to increase output and market share. They did so, however, only after they had developed production, distribution, and research facilities and skills comparable to those of the four leaders.

Comparable patents, exchange of technical information, and market allocating agreements were significant in the chemical industries. There global cartels of a small number of large firms were relatively effective in glass, rayon, explosives, synthetic nitrates, dyes, pharmaceuticals, and films.[31] In the most technologically advanced organic and inorganic chemical industries, the Germans who had built the most efficient plants and created the most effective worldwide marketing organizations continued to have a major say in the drawing up of agreements and continued to have the economic power to enforce them. The Americans, because of the antitrust laws, rarely joined agreements to set price and output; but they did, through patent and licensing arrangements, agree to stay out of Europe if European firms, with some exceptions, stayed out of the U.S. markets. As in machinery, research and development continued to be done at home.

In the production of metals, both ferrous and nonferrous (group 33) the massive throughput needed to maintain minimum efficient scale kept the leading enterprises highly centralized. These firms rarely found it economical to build plants abroad, even though tariff barriers and transportation costs were high. The construction of costly plants in foreign economies at such a scale usually so expanded local capacity that existing demand was not enough to maintain a throughput necessary to achieve a competitive advantage. These firms, however, had branch offices abroad and employed a small professional international sales force to obtain orders with precise specifications to be delivered on a precise schedule.

Because this highly concentrated production in these industries with cyclical demand often brought the threat of global overcapacity, contractual cooperation by means of formal cartels became prevalent. However, cartels in iron and steel were rarely successful, even though in Europe they had the support of their national governments and even though the U.S. producers had withdrawn from the European markets.[32] The number of firms may have been too numerous and the products too differentiated to enforce effectively the agreements reached. On the other hand, in nonferrous metals—copper, nickel, and aluminum—such contractual agreements were more successful and longer lived. Far fewer firms were involved. In addition, the enormous scale economies generated by electrolytic refining technologies provided the pioneers with the economic power to enforce agreements, as their unit costs usually remained well below their somewhat smaller competitors.

In these ways, the strategies and structures of multinational companies competing for shares of the global markets reflected during the interwar years the cost advantages of scale, the pace of technological change, the existing restrictions to trade, and the contractual arrangements between a small number of competitors to set price and output and to allocate markets. In some cases, scale economies inherent in different technologies of production encouraged centralization, in others decentralization of decision making within multinational enterprises. Tariff barriers often hampered the rational exploitation of scale economies, led to building smaller plants for national markets, and often further decentralized decision making. On the other hand cartel arrangements usually centralized decisions as to price, output, and markets.

THE POST–WORLD WAR II ENVIRONMENT: MARKETS, TECHNOLOGY, AND MANAGEMENT

The underlying institutional framework within which global competition was carried on, created in the last decades of the nineteenth century and firmly established during the interwar years, has remained intact until today. Even the trauma of World War II did not alter basic patterns of international business. U.S. firms, of course, came out of the war with an enhanced position. Nevertheless, the devastated European and Japanese economies recovered with impressive speed. In both areas, the same large integrated multinational enterprises continued to dominate in the same industries for much the same reason. Even the German chemical and machinery firms, whose facilities were often totally destroyed and personnel scattered and whose homeland was divided, soon had positions in national and global markets almost as strong as they had been in the 1930s.

Indeed, in the postwar global setting there was more of a change in the number and activities of the nation states than in those of the multinational enterprises. While the United States and the countries of Western Europe continued to dominate world trade, with Japan as the one new challenger, a host of new nations, many former colonies or dependencies of the dismembered empires, appeared. Nearly all hoped to industrialize and to achieve economic growth by adopting the new technologies of production and distribution. Moreover, the world was quickly divided between the East and West—the Communist and non-Communist—with very little trade between the two. The key actors remained much the same, but the environment in which they operated changed swiftly. This transformation rested, as it had in the past, on expanding markets and continuing technological change.

Changing Markets

During the two decades after 1950, both domestic and international markets expanded at a faster rate than at any time since the 1880s.

Not only did population increase, but of most importance, so did per capita output and income.[33] The continuing rapid migration from the country to the city rapidly increased the numbers of customers for processed and manufactured products.

During these same decades of U.S. hegemony, traffic and other barriers that divided markets melted away. When trade restrictions lingered, they were often the result of legislation to encourage economic growth and health. In the 1950s and 1960s, the economic autocracy of the interwar years all but disappeared. The formation of the European Community (EC) by the Treaty of Rome in 1957 created a domestic market for European manufacturers that was large enough and growing fast enough to permit scale economies in production comparable to those that had long existed in the United States.

The decline in trade barriers was accompanied by a breakdown of prewar cartels and other interfirm contractual agreements. Free trade and competition, basic tenants of the U.S. economic creed, were exported abroad. The victorious U.S. authorities dismantled the cartels in Germany and dismembered the *zaibatsu* in Japan. Then, in 1956, the British Parliament passed the Restrictive Practices Act, the first Parliamentary legislation to challenge the legal and economic validity of maintaining market power through contractual cooperation.[34] Soon the EC began to develop a body of legislation that did much the same for the new broad continental market.

Even earlier, during and immediately after the war, the U.S. Justice Department had moved successfully against U.S. firms that through patent and license agreements had been able by legal means to allocate markets abroad and otherwise maintain their position in international trade. The Department court rulings prohibited the making of such agreements even with foreign affiliates in which they held a substantial, often controlling share (but less than 50 percent), of the voting stock.[35] These rulings at home, plus the new legislation abroad, weakened the position of U.S. machinery makers, particularly the manufacturers of electrical equipment in foreign markets. On the other hand, they opened the European market to the U.S. chemical companies. Thus, by the end of the 1950s, competitors in the enlarged and rapidly growing global markets no longer relied on contractual cooperation to maintain market power. Competitors in nearly all the global oligopolies now competed functionally and strategically for market share.

Technological Change

As important as expanding markets in shaping the global competitive environment was the diffusion of existing technologies and the development of new ones, particularly the new communications and information technology. The rapid transfer of highly product-specific production technologies during and after World War II eroded U.S. strength in international markets in light machinery, automobiles, and other vehicles and the German position in the technologically complex industrial, chemical, and machinery markets. Especially striking was the loss of U.S. dominance in the production of automobiles and commercial vehicles. Although Morris and Austin in Britain and Renault and Citroën in France had adopted U.S. production techniques during the interwar years, they had by 1939 gained only a tiny share of the international markets.[36] In Italy and Germany, such methods first appeared with the opening of the large Fiat plant in 1936 and the Volkswagen works in 1940, too late to affect prewar global competition. After the war, such companies quickly expanded output to meet domestic and then European demand. For example, by 1960, Volkswagen and Mercedes had joined Ford and General Motors's Opel as the dominant German firms, the four accounting for 87 percent of the nation's automobile output. In the next decade, the United States rapidly lost market shares to new competitors abroad and then in the 1970s at home. To a somewhat lesser extent, the same pattern occurred in the production of sewing, office and agricultural machinery, and consumer appliances. In electrical machinery and other volume-produced industrial equipment some of the major firms voluntarily, though temporarily, withdrew from the European markets partly because of the Justice Department's antitrust drive.

Far more dramatic was the new competition from Japan. As Hirschmeier and Yui noted in their *Development of Japanese Business 1969–1973:*

> Between 1950 and 1967 Japanese industry set records of technological transfers. A total of 4,135 licenses were purchased by Japanese industry, mainly from the USA, over half in the field of machinery construction and about 20% in the field of chemical industries. During the same time exports of licenses amounted to only about 1% of the money spent on imports of patents and licenses.[37]

This unprecedented transfer of technology and the rapid growth of the Japanese domestic market permitted that nation to adopt for the first time the new processes of mass production and mass distribution that have been at the center of Western industrial economies since the turn of the century. As a result, the Japanese firms, particularly in mass-produced automobiles and appliances, moved into existing global oligopolies. For example, the Japanese market, which absorbed less than 500,000 new cars in 1964, took 1.13 million in 1967, 2.38 in 1970, 2.93 in 1973, and 3.10 million in 1979.[38] Where in 1960 there were 5 passenger cars in Japan for every 1,000 inhabitants, by 1973 there were 185 passenger cars per 1,000. This expanding market in turn permitted scale economies based on the newest production technologies to reduce costs enough so that the large Japanese firms (by 1974 there were only five major manufacturers) were able to move decisively into world markets. Their great success was in mass-produced consumer durables where managerial efficiencies in manufacturing and distribution were particularly pronounced. In branded packaged food and consumer products, in chemicals, oil, and even rubber the Japanese challenge was less formidable.

The coming of new technologies may have had an even more direct impact on global competition than did the rapid transfer of existing ones. Of paramount importance were the innovations in the compilation, collation, and communication of business and industrial information. If the revolution in transportation and communications of the mid-nineteenth century created modern global competition, the second communications revolution of the mid-twentieth century has probably been the major force in shaping its current configuration. The railroad, supplemented by the truck, and the ship still carry the bulk of the world's goods. (The substitution of oil for coal as an energy source has not significantly changed the speed or regularity of land and sea transportation.) But in communications the jet plane, the new telecommunications technology, and the computer have achieved in Raymond Vernon's words, "The spectacular shrinkage of space."[39] The coming of the teleprinter made cable communication more effective. In 1956, the first trans-Atlantic telephone cable went into operation, greatly improving the speed and reliability of overseas telephone communication, which since 1927 had depended on radio transmission. In 1965, global communication was further improved when the first commercial communications satellite went into operation. In the 1950s, the introduction of the jet airplane for commercial travel greatly expanded the possibility of face-to-face communication. By 1960, international commercial flights already logged 26 million passenger miles. In 1974, the figure had reached 152 billion.

Even more impressive has been integration of the computer into the new international communications networks. In the late 1950s, the computer was being rapidly adapted for a multitude of business purposes. Then, in 1963, Honeywell demonstrated its potential for international coordination by placing a terminal in Britain to control a computer at a plant in Massachusetts with control signals being sent by the standard Telex line.[40] At first the new transborder computer networks were used largely by service companies for hotel and airline reservations, stock market quotations, and banking and insurance transactions. By the 1970s, multinational industrials were, according to one authority, increasingly relying on such computer data, usually through leased lines, "to coordinate production and marketing, to coordinate financial management, to share data-processing resources, to reduce costs of telephone circuits; to share scientific and technical research; [and] to improve accuracy and security of information transfer (e.g., by using standard message formats and data encryption techniques)." By the mid-1970s, such international computer communication was becoming an integral part of the control and information systems of British and European multinational enterprises, though possibly less so for their Japanese counterparts.[41]

The Transfer of Managerial Methods

The 1960s and early 1970s were not only the years of technological innovation and diffusion, but they were also those of the diffusion of management techniques. The United States had pioneered in the development of administrative and financial controls and overall management structures essential to the effective operation of a giant multinational enterprise. This was because, in part at least, U.S. firms competed more aggressively, both functionally and strategically, for market share than did those in Europe. It was these administrative skills and procedures that J. J. Servan-Schreiber had in mind when he wrote *The American Challenge*. The opening sentence of that book published in 1968 reads, "Fifteen years from now it is quite possible that the world's third greatest industrial power, just after the United States and Russia, will not be Europe, but American industry in Europe. Already, in the ninth year of the Common Market, this European market is basically American in organization." For Servan-Schreiber, the U.S. challenge was not one of U.S. financial power, nor the European "technological gap." "On the contrary, it is something quite new and considerably more serious—the extension to Europe of an *organization* that is still a mystery to us."[42]

It did not remain a mystery long. When Servan-Schreiber wrote, the transfer of techniques was already well under way. By 1970, for example, over half of the largest one hundred industrials in Britain had used the services of a single U.S. firm, McKinsey & Company, to reorganize their management structures. In most cases such reorganizations meant a transformation of a holding company into a modern multidivisional enterprise by rationalizing the operations of subsidiaries and by building a large corporate office.[43] In Germany, the growth and diversification of large firms brought to a lesser extent the adaptation of the similar structure; while in Japan came a rapid adaptation of quality, inventory, and other U.S. control systems as well as the importation and improvement of overall management structures worked out in the United States.

In the 1970s, then, the global environment in which the United States had dominated in the 1950s and 1960s was rapidly transforming. The two major sets of actors remained the same as they had been for almost a century. There were more nation states, but only one, Japan, had acquired industrial strength comparable to that of the United States and Western Europe. Nor was there a large turnover among the multinationals. They still clustered in much the same type of industries, and in most of these industries the leaders in the first decade of the century were still the leaders in the ninth decade. What makes Servan-Schreiber's prophecy so ironic today is that continuing changes in markets, technology, and managerial methods transformed the environment in which the multinationals competed and eroded the historical competitive advantages of U.S. firms.

CONCLUSION: THE CHALLENGE OF THE NEW COMPETITION

Because of these changes, the number of global competitors in the industries dominated by large enterprises increased and the competition between these players became based much more on functional and strategic competition than was true earlier. The U.S. firms lost their technological edge in many industries. At the same time they lost their managerial advantage. These losses were particularly serious in the mass-produced machinery industries—automobiles, appliances, and office, sewing, agricultural, and other light machinery. The leaders in these industries failed to maintain the functional and strategic competitiveness that won them global superiority early in the century. On the other hand, oil and rubber companies that continued to think in global terms and the chemical and pharmaceutical firms that further honed their functional and strategic skills as they made their initial moves into European markets were less adversely affected by the new competition.

At the very moment the United States was losing its technological and managerial advantages, the new communications and information technologies were permitting managerial teams to coordinate the flows essential to exploit fully the scale economies of global markets. The newcomers, with a smaller investment in older plants and equipment and with less of a commitment to specific established administrative practices and procedures, were able to move more quickly to exploit the cost-reducing possibilities of the new telecommunications and computer technologies. The United States was paying the price of early success. It faced a new environment with a huge investment in increasingly outmoded capital equipment and with inevitable institutional rigidities resulting from decades of successful competition under different conditions.

Moreover, the challenges of this new environment came at a time when the United States relied far more on global markets to provide the volume necessary for profitable operations than it had earlier in its history. For three decades after 1940, U.S. industrial enterprises had steadily increased their working force. The unprecedented war demands were followed by almost a quarter of a century of the greatest expansion of the domestic market ever witnessed. When the growth of the U.S. economy began to slow down in the early 1970s, U.S. firms, in order to maintain scale economies, responded by expanding rather than contracting their overseas activities. The direct investment by U.S. enterprises abroad rose from $75.5 billion in 1970 to $227.3 billion in 1981.[44] At the same time, foreign direct investment in the United States rose from $13.3 billion to $89.8 billion. Thus, as global markets have become increasingly vital to the United States, so U.S. markets have become more important to foreigners. Such global competition has become increasingly based on functional and strategic efficiencies in situations in which the U.S. firms no longer have a technological or managerial edge. The resulting challenge is the greatest challenge the modern U.S. industrial enterprises have had to face since the beginning of this institution a century ago.

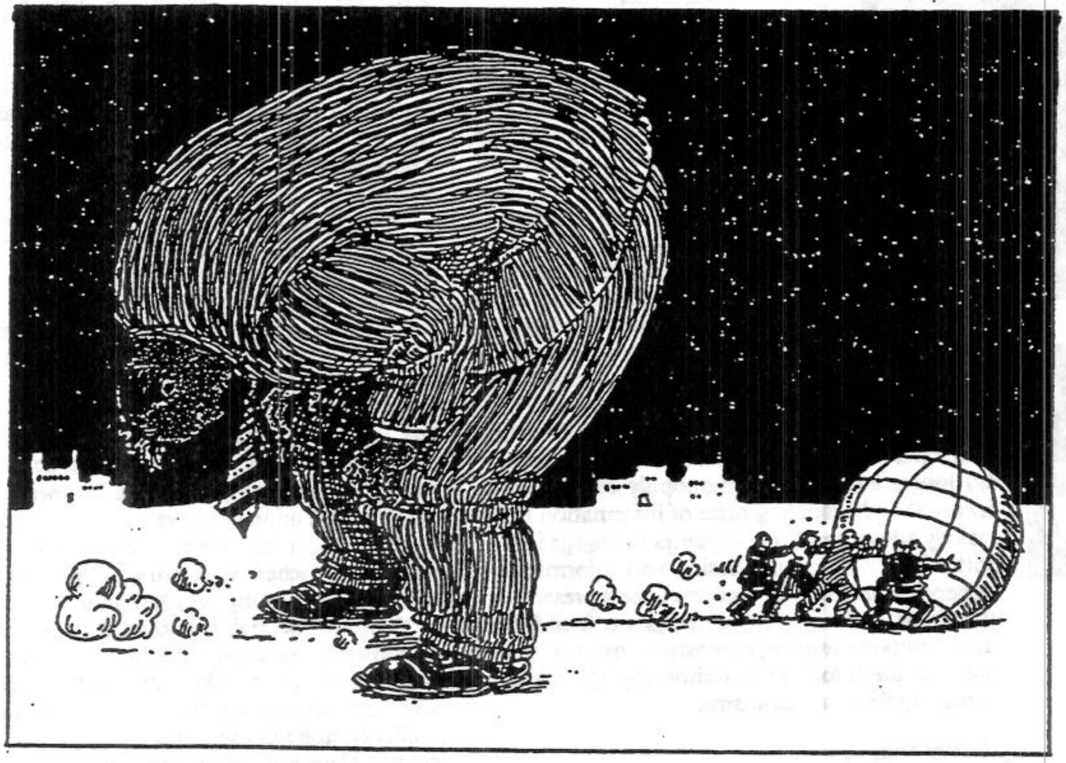
This review has touched only on the major factors that helped to shape the modern global competition since it came into being in the late nineteenth century. It has only hinted at the variations and the subtleties in the constantly changing pattern. In each industry, the multinational enterprise operated with somewhat different technologies of production and different methods of distribution. Each nation state, as a result of its past history and current economic situation, has faced different issues and challenges in working out relationships with both home-based and foreign multinationals.

NOTES

1. A. D. Chandler, Jr., *The Visible Hand: The Managerial Revolution in American Business* (Cambridge, Mass.: Harvard University Press, 1977), 230–32.
2. Details and documentation are given in a case by A. D. Chandler, Jr., on "The Standard Oil Company: Combination, Consolidation and Integration," in A. D Chandler, Jr. and R. S. Tedlow, *The Coming of Managerial Capitalism: A Case Book in the History of American Economic Institutions* (Homewood, Ill.: Richard D. Irwin, 1985).
3. B. W. E. Alford, *W. D. & H. O. Wills and the Development of the U.K. Tobacco Industry* (London: Methuen, 1973), 143–49. Also Chandler, *The Visible Hand*, 249–58. For similar new high-volume technologies in the production of other branded packaged products see Chandler, chap. 9.
4. S. Kaku, "The Development and Structure of the German Coal-Tar Dyestuffs Firms," A. Okochi and H. Uchida, eds., *Development and Diffusion of Technology* (Tokyo: University of Tokyo Press, 1979), 78.
5. This statement is based on a review of histories of and annual reports and pamphlets by the leading rubber companies.
6. H. Livesay, *Andrew Carnegie and the Rise of Big Business* (Boston: Little, Brown, 1975), 102–6, 155. When in 1873 Carnegie opened the first works directed entirely to producing rails by the Bessemer process he dropped cost to $56.64 a ton. By 1889, with increase in sales the costs fell to $25 a ton.

[illegible]

MULTINATIONALS



WHARTON REPROGRAPHICS

Who wants to be a giant?

Multinational companies have been subjected to ferocious, often humbling competition in recent years. All to the good: it has made them slim down and spruce up, argues Adrian Wooldridge

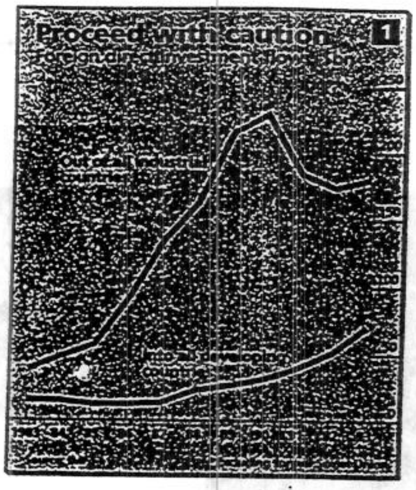
MULTINATIONAL companies have been accused of many things in their time: representing the unacceptable face of capitalism; making the world a less colourful place by promoting bland homogeneity; even doing the devil's work on earth. But what has really hurt the men and (handful of) women who run the world's multinationals is a new criticism that has emerged in recent years: that they are failures.

The evidence is embarrassingly abundant. The list of companies that ended the 1980s in worse shape than they started reads like a Who's Who of leading multinationals: Boeing, Caterpillar, Dayton-Hudson, Du Pont, Salomon Brothers, Texas Instruments, Westinghouse, Xerox... Huge losses were run up by household names: $457m in 1991 by Citicorp, America's biggest international bank; $23.5 billion in 1992 by General Motors, which more or less invented modern corporate management; and $8.1 billion in 1993 by IBM, the company that had dominated every national market it entered. Nor was the blight confined to the United States. Philips, a Dutch electronics giant, got into deep trouble, and Japan's Matsushita and Germany's Daimler-Benz had to undergo drastic surgery.

The business heroes of the recent past have not been multinational company men, but entrepreneurs who have turned start-ups into money machines: Bill Gates of Microsoft, Steve Jobs of Apple, Richard Branson of Virgin. Anybody with a bright idea and a rented garage, it seemed, could take on the giants and win. Peter Drucker, a veteran management guru who started his career writing panegyrics to General Motors, announced that "The 'Fortune 500' [the annual listing of America's biggest industrial companies] is over." Another guru, Tom Peters, who once regarded IBM and other American giants as paragons of excellence, now celebrates the "demise of size".

Yet on the face of it, prospects for the world's giant corporations were looking better than ever. Governments were progressively lowering trade barriers. Restrictions on foreign investment were being removed. Companies were shifting formidable quantities of money from country to country with almost unprecedented ease: since 1983 foreign direct investment has grown five times faster than world trade and ten times faster than world output, though that growth has eased in recent years (see chart 1 on the next page). Governments everywhere regarded multinationals not as predators to be avoided but as sources of investment, advice and above all jobs. Even the Chinese government, which had once promised to wipe these running dogs of capitalism from the face of the earth, started inviting them to build factories instead.

On the face of it, too, the increasing globalisation of business seemed likely to play to the multinationals' strengths. They would be able to manufacture their goods wherever they could find the



Proceed with caution 1

best combination of price and quality, and distribute them wherever they could discover (or create) a demand. Many pundits predicted that a handful of global firms would soon carve up the world between them.

As it turned out, globalisation exposed the multinationals' weaknesses. Deregulation and lower trade barriers have reduced the value of their carefully cultivated relationships with governments and expensively accumulated knowledge of local regulatory quirks. It is no longer necessary to be a classical multinational to penetrate global markets, as companies as diverse as Microsoft and Swatch have demonstrated. The spread of modern management techniques means that multinationals no longer have a monopoly of managerial wisdom. An epidemic of pirating, reverse engineering and other forms of industrial theft means that they no longer even enjoy a monopoly of their own ideas. The plummeting price of information technology allows smaller companies to engage in the sort of information processing and information-dependent innovation that was once a preserve of the giants. Indeed, many of the giants found that their computer systems, expensively installed at the dawn of the information technology age, had become hopeless anachronisms.

Heavy baggage

The trouble with many multinationals is that they are legacies of a very different era. Many grew up in the heyday of command-and-control management, when strategy was made by a tiny elite at the top, work was broken down into its simplest component parts and workers were monitored by layer upon layer of managers. But today fashion is so fickle and markets so quicksilver that decisions are best taken by front-line workers rather than by lethargic middle managers.

Above all, most multinationals are a product of the doctrine of "economies of scale"—the idea that producing things in ever larger volumes brings down unit costs, so that big companies will do better than small ones. But economies of scale have been declining as techniques such as lean manufacturing and just-in-time production have shifted the emphasis from size to timeliness, and computerised production has allowed factories to produce customised goods at mass-produced prices.

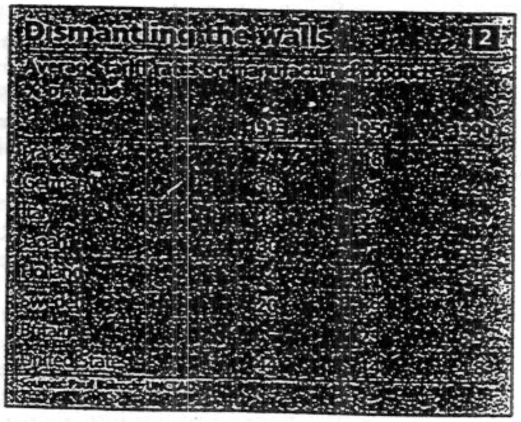
At the same time, the diseconomies of scale are looming ever larger. Giant companies generate bureaucratic bloat; giant factories create shopfloor alienation; and giant corporations often fail to attract creative workers, or to make good use of those they have. The average size of workplace has been falling steadily throughout the industrialised world since the late 1960s as large organisations try to break themselves up into smaller parts. Worse still, the standardised products which pour out of large factories are suffering from a double squeeze: fashion-conscious customers increasingly want goods tailored to their particular needs, and value-conscious customers are less willing to pay a premium price for a global brand.

Many multinationals are also legacies of an age of high tariffs and high transport costs. Companies such as Ford and Unilever essentially set up clones of themselves in all the countries where they operated, complete with head offices, engineering and design facilities and production plants. This was expensive and wasteful at the best of times. At one point Ford had two separately built and designed Escort models on the road. Sometimes different national subsidiaries of the same firm were locked in head-to-head conflict. But as trade barriers fell (see table 2), this strategy proved unsustainable.

Not all big firms trod this multidomestic road. The Japanese companies, which appeared on the international scene in the 1950s when tariff barriers were falling, preferred to produce as much as possible at home for export. But western governments eventually objected to the huge trade surpluses this created—especially when the Japanese seemed so adept at keeping foreigners out of Japanese markets. Even when Japanese companies started moving production abroad, westerners complained that they were simply exporting "screwdriver factories" which added little value. As the 1980s wore on, both the multidomestic and the export-oriented approaches to running large companies seemed to be running into the buffers.

Yet multinationals are not on their way out. On the contrary, this survey will argue that a period of cut-throat competition has left them fitter than they have been for decades. They are mimicking their smaller competitors by shrinking their head offices, removing layers of bureaucracy and breaking themselves up into constellations of profit centres. They are learning to combine economies of scale in product development with sensitivity to local tastes. And they are beginning to use their ability to operate across borders to much better effect.

Above all, multinationals have changed their ideas about where their competitive advantage lies. They used to think that their most precious resource was capital, and that the prime task of management was to allocate it in the most productive way. Now they have become convinced that their most precious resource is knowledge, and that the prime task of management is to ensure that their knowledge is generated as widely and used as efficiently as possible. The next article will look at the effect of this change on the way they are being managed.



Dismantling the walls 2

The world turned upside down

OVER the past decade multinational companies have been re-examining almost every item of received managerial wisdom. Once they prided themselves on their "vertical integration"—running themselves as self-contained empires. At one point Ford even smelted its own steel. Now they are "sticking to their knitting"—concentrating on their core business and contracting out everything else. Even the world's oldest multinational, the Roman Catholic church, has contracted out the management of its worldwide data network to General Electric. Once multinationals devoted enormous resources to gathering information, getting it processed by legions of managers and handing it to a strategy-making elite. Now they are putting a computer on every desk and giving power to front-line workers. Once they prided themselves on their vast headquarters and workforces to match. Now they are miniaturising their head offices and cutting their payrolls to the bone.

This change in thinking has had dramatic consequences. In 1993 big American companies announced nearly 600,000 lay-offs, 25% more than in 1992 and 10% more than in the depths of the recession in 1991. Even paternalistic companies such as IBM, which used to pride themselves on offering a job for life, have been forced to join in the blood-letting. And this time the burden has fallen not just on the hapless workers. Companies have been cutting their headquarters staff and getting rid of entire layers of management.

When Jack Smith took over as boss of General Motors in 1992, he cut the corporate bureaucracy from 13,000 to 2,000 and moved his office from that symbol of corporate America, the 14th floor of GM's headquarters in Detroit, to a new technical centre in the suburbs 15 miles away. Ford has reduced the number of its management levels from ten to seven. Percy Barnevik, who heads Asea Brown Boveri (ABB), a Swiss-Swedish engineering giant, applies a "30% rule" to businesses he acquires, allocating 30% of central staff to independent profit centres, 30% to operating companies and getting rid of 30% entirely; only 10% are left in head office. His corporate headquarters in Zurich has a staff of about 170.

A touch of anorexia?

All this has brought both institutional and personal disruption in its wake. That disruption is set to continue, despite recent fears that the slimming might have gone too far—some talk of "corporate anorexia". Even the Japanese are being forced to rethink their commitment to lifetime employment. This is not necessarily a bad thing. Many middle managers lived boringly industrious lives without adding any real value. Contracting-out some types of work has created career structures where none existed before: computer programmers, for example, are far better off working for a specialist such as EDS than they were as obscure number-crunchers in a sausage-making factory.

Large companies have been trying to imitate the virtues of their smaller rivals by breaking themselves up into smaller units and giving them plenty of freedom to do things their way. For example, Bertelsmann, a German publisher, has created over 150 profit centres. ABB has subdivided itself into 65 business areas, 1,100 local companies and 4,500 profit centres. Cap Gemini, a French computer software and services firm, operates "free-standing branches", each organised as a small local company of 40-100 people. The branches can be combined for specific tasks, and are grouped geographically for fiscal and tax purposes. Canon, a Japanese office-equipment maker, has turned the heads of its numerous operating units into mini-chief executives, with full responsibility for growth and profitability. Matsushita has created a "one product, one division" structure.

It all comes down to a change in thinking about how companies use knowledge. The old view was that strategy should be set by a tiny elite at the top, and translated into action by workers who were simply interchangeable parts. "How come when I want a pair of hands I get a human being as well?" Henry Ford once famously asked. Now the men at corporate headquarters realise that decisions are often best taken by those who spend their lives developing products or dealing with customers.

At the same time companies have downgraded many national head offices and tried to get them to operate as part of a global system. Ford abolished its separate national units at the beginning of this year. Louis Gerstner, chairman of IBM, is reducing the power of his company's country heads and creating 14 broad groups that can deal with global customers on a global basis. Even Unilever, which once prided itself on operating as a local company wherever in the world it went, is starting to rationalise in Europe.

Most multinationals have been shifting power to the regional level, particularly in Europe. General Motors and Ford established regional headquarters in Europe long ago—although Ford recently reversed the process and reorganised itself by worldwide product group. A number of companies have been giving subsidiaries responsibility for global products or global functions, partly to disperse decision-making throughout the organisation and partly to capture local expertise. IBM, AT&T and Hewlett-Packard have moved the worldwide headquarters for some products from America to Europe. Glaxo, Unilever, Siemens and Hoechst have all created global product groups. Nestlé has moved the headquarters of its pasta business to Italy. Johnson & Johnson has given its Belgian subsidiary world-wide responsibility for pharmaceuticals.

As if inventing new products and penetrating new markets was not hard enough, the multinationals are having to rewrite their management manuals as well

Slimmer and fitter? 4

Selected companies' headcount reduction in 199[illegible]

Reduced by <10%		Reduced by 10% or more	
[illegible]	[illegible]	[illegible]	30
[illegible]	[illegible]	[illegible]	17
[illegible]	[illegible]	[illegible]	17
[illegible]	[illegible]	Eastman Kodak	17
[illegible]	[illegible]	Owens-Illinois	16
[illegible]	[illegible]	Union Carbide	13
[illegible]	[illegible]	[illegible]	13
[illegible]	[illegible]	[illegible]	11

The United Nations Conference on Trade and Development, which keeps a permanent watch on what it dubs "transnational corporations", distinguishes between simple and complex integration. Simple integration means that companies contract-out routine production to the developing world but keep their most sophisticated operations in the home country. Nike, an American sportswear maker, keeps all its product development and marketing in Beaverton, Oregon, its home town, but sub-contracts production to 40 different locations, mostly in South and South-East Asia. If wages in one host country rise, the firm simply shifts production. In complex integration, companies locate all their activities according to the logic of the market, and disperse decision-making throughout the organisation. Their hallmark is the endless flow of information in all directions instead of a command-and-control system.

Xerox, an office-equipment maker, has been particularly successful at moving from a multidomestic structure to complex integration. In the early 1980s Xerox's various national affiliates, such as Britain's Rank Xerox and Japan's Fuji Xerox, did everything from design to distribution on their own. But competition from cheaper, more flexible Japanese companies such as Canon forced them to integrate their activities. The company standardised its operations across borders, introduced global product-development teams and centralised procurement, reducing the number of suppliers from 5,000 to 400.

All these changes have required formidable political skills. Getting senior managers to relinquish power is hard enough; making them co-operate in their own downgrading is often impossible. Companies have had to deal with wholesale revolts by corporate barons and silent, sullen sabotage from within. Both Jack Welch at General Electric and Sir Colin Marshall at British Airways found that they had to start by sacking senior people.

The best managers have gone out of their way to explain why change is inevitable. Ford created a rolling programme to explain why it was pursuing a global strategy throughout the organisation. Its headquarters in Dearborn, Michigan, flaunts a huge board underlining all its senior managers' commitment to the programme. When Gilbert Amelio took over National Semiconductor, an ailing Californian chip company, in 1992, he invited his 2,000 top managers to California to spend a week contemplating the company's predicament and planning its future. As well as addressing every class, he visited all the company's sites around the world to answer his employees' questions and explain why things had to change.

The best managers also make concessions to win employees' hearts and minds. When Mr Barnevik merged ASEA, his Swedish engineering group, with Brown Boveri, a Swiss company in the same business, he established his headquarters in Zurich rather than Stockholm to make the operation look less like a takeover. For similar reasons, when Nestlé took over Rowntree it put the headquarters of its confectionery business into Britain.

All this has involved formidable organisational changes too. Companies have had to establish communication channels where none existed, and find new uses for redundant parts of the organisation. This has created work for an army of consultants and generated a plague of management fads.

These twin trends towards flatter hierarchies and closer integration across borders have pushed multinationals towards a new model which they like to call a "networked organisation" and in which activities and decision-making are widely dispersed. But even as they transform themselves into globally integrated networks, companies must beware of losing their local roots.

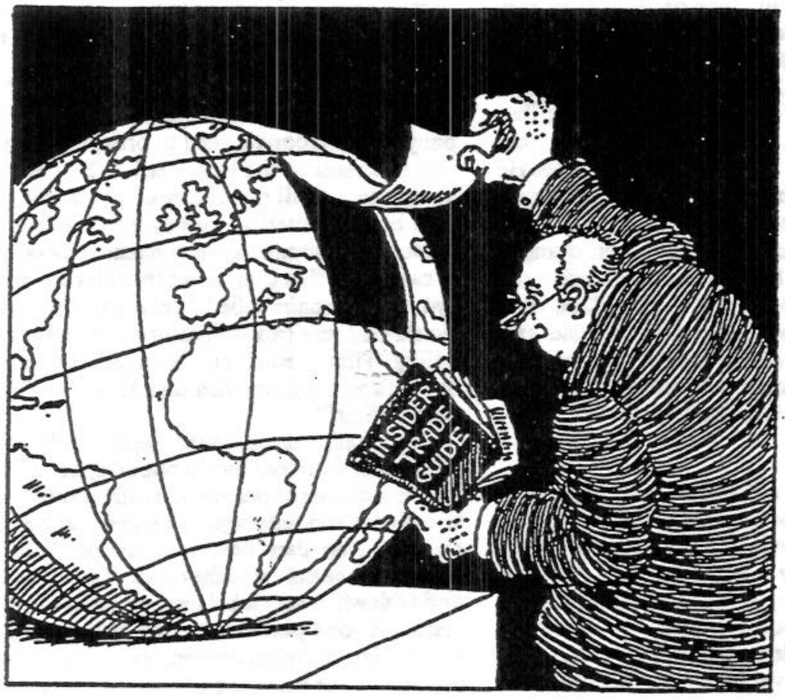
Insider trading

Multinationals need to become local insiders in lots of different places

MENTION the word "multinational", and most people think of borderless mobility—of companies at home everywhere and nowhere, moving huge quantities of men, money and materials around the globe in the restless pursuit of profit. In 1993, despite the continuing recession, multinationals invested some $232 billion the world over; by 2020 these flows may well have increased fourfold. Aurelio Peccei, a director of Fiat, once claimed that the multinational corporation was "the most powerful agent for the internationalisation of society." Carl Gerstacker, sometime chairman of Dow Chemicals, confessed that he had "long dreamed of buying an island owned by no nation and of establishing the world headquarters of the Dow company on the truly neutral ground of such an island, beholden to no nation or society."

Mention of the word "multinational" also makes people think of global products—of consumers in Greenwich Village and some village in Bengal drinking identical soft drinks. Theodore Levitt, a marketing guru at Harvard Business School, once argued that companies no longer need to be so "respectful" of local quirks and peculiarities, and that global companies can sell the same thing in the same way anywhere.

Multinationals should beware of following Mr Levitt down this path. After a brief flirtation with



globalisation, companies such as Nestlé and Unilever now realise that their local managers represent an invaluable resource. Electronics companies such as Motorola and Compaq, which started off by organising themselves along global product lines, are adding a national dimension. General Electric has broken with its strict practice of dividing its operations into global product lines by setting up a regional headquarters in Asia.

Even the most powerful global brands have had to bow before local differences. People in the south of Japan like their Coca-Cola slightly sweeter than people in Tokyo, and the company obliges. PepsiCo was puzzled why one of its best-selling products, "7-Up" remained on the shelves in Shanghai until it discovered that, in the local dialect, the phrase means "death through drinking". Even those pillars of American puritanism, IBM and Disney, have dropped their strict no-alcohol policy in France. Philip Morris has had to make local adjustments to its familiar advertising symbol, Marlboro man: in Hong Kong the advertisement focuses on the horse, because the man reminds locals of a coolie, and in Argentina the man was dropped entirely for a while, because cowboys were regarded as low-class wasters. Companies are also learning to their cost that the apparent convergence between different cultures has not gone as far as they thought. EuroDisney failed to take off, among other things, because the company assumed that Mickey Mouse and other cartoon characters would be as familiar in Europe as they are in the United States, and did not invest nearly enough in promoting their product.

In praise of roots

Companies have also started to remember that their local roots can be very valuable commodities. Michael Porter, a management guru at Harvard Business School, has pointed out at length that different countries and regions have different competitive strengths: the Germans excel at high-quality engineering and chemicals, the Japanese at miniaturisation and electronics, the British at pop music and publishing, the Americans at films and computers. If they are to retain those advantages, they need to remain in constant contact with testing competitors and world-class suppliers. Becoming more global may be no help in that.

Even when they do venture out of their home territories, many multinationals move no further than their local regions. Foreign investment tends to cluster around the investor's home country: American firms put money into Central and South America, the Japanese into a handful of Asian countries, and European firms into other European countries, including Eastern Europe. The attractions include ties of language or culture and the closer contact made possible by shorter travelling times. The Japanese, for example, find the informal alliances and extended company networks of their neighbours in South-East Asia much more congenial than the Europeans do. Some modern management techniques reinforce this trend towards regionalisation: the fashion for keeping stocks low (lean manufacturing) and delivering supplies only when they are needed (just-in-time production) means that suppliers need to be reasonably close to their customers.

But the main reason why companies need to cultivate their local roots is that this is where some of their most precious resources lie. Alfred Marshall, one of the founders of modern economics, once concluded in an analysis of Sheffield steelmaking that in some regions certain skills seemed to be "in the air". The education system fosters them, the local infrastructure and political system helps them along and competition from a cluster of similar companies encourages an informal change of ideas.

Some of these centres of excellence are familiar: Silicon Valley in computing, Hollywood in film-making, the Prato region of northern Italy in fashion and design. Others are less predictable. Procter & Gamble, a packaged-goods company, recently asked people from all round the world to come up with ideas for an advertising campaign, and was surprised to find that the clear winner came from Taiwan. But all these centres of excellence have one thing in common: they rely on an unusual combination of natural and man-made factors, and they are sustained by distinctive local traditions and cultures. In an increasingly mobile world, they are stubbornly immobile resources.

According to John Stopford, a professor at the London Business School, one of the most popular ways to become a fully-fledged insider is for companies to give worldwide responsibility for particular products to affiliates in the right areas. Thus Du Pont has moved its electronics-related businesses to Japan, Siemens has shifted its air-traffic management to Britain, and Hyundai has located its personal-computer business in the United States. Another way is to form joint ventures or alliances with companies already in the right place. Being an insider can also help companies to master and manipulate local regulatory systems. And in many countries it is vital for getting your products distributed. For example, the Americans have had prob-

lems selling their cars in Japan partly because they have paid no attention to a distribution system so personal it uses door-to-door salesmen.

McDonald's and Coca-Cola, on the other hand, have done well out of working with local distribution systems in Japan. McDonald's allowed a local entrepreneur to adapt its product to local tastes. He set up small stores in central places such as Tokyo's Ginza, rather than the large suburban stores favoured by the parent company, and even changed Ronald McDonald to Donald to accommodate the local difficulty with the letter "r". Coca-Cola went further, adjusting its formula to suit local tastes and launching a stream of local products. Coca-Cola was the first foreign company to join the Japan Federation of Employers' Associations in 1975.

Novelty value

Joseph Raudabaugh of A.T. Kearney, a consultancy, points to a more subtle reason why companies need to be insiders, particularly in mature markets such as Japan. Many companies find that the longer they stay in Japan, the less profitable they become: their fixed costs go up inexorably but their products sell less and less well, because local companies make cheaper versions and because local consumers crave novelty. The only way to escape is to start developing products locally. Two-thirds of the products Coca-Cola sells in Japan are made specifically for the Japanese market.

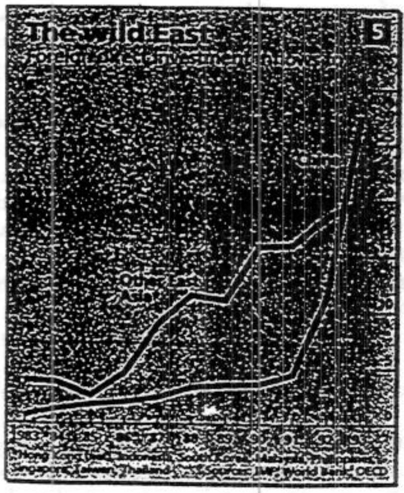
But there are strict limits to this "insider" strategy. It is not enough to be just another local company: multinationals need to bring new products and processes to the local market. Besides, many people get a thrill out of consuming foreign products, be they American soft drinks or British suits. Some local companies even pretend to be outsiders to cash in on that quirk. Twenty miles south of Tokyo, a store-chain called Nichii has established a would-be American mall complete with boutiques selling "Fifth Avenue" clothes, a gadget store called So-Ho, and a gigantic fish tank known as "Fisherman's Wharf".

Above all, multinationals need to make a virtue of being insiders not just in one area, but in many. They need to act as dealers in locally-rooted insider knowledge and arbitrageurs between centres of excellence. And they need to combine these local skills with the expertise they possess as a company, and with whatever global resources they can get access to. A good place to discover how to strike the appropriate balance between being global and being local is the emerging Chinese market, the subject of the next article.

Managing on the frontier

Other countries, other mores

SINCE the late 1980s, multinational managers have been forced to rewrite their textbooks not once but a dozen times. As well as reinventing their organisations back home, they have had to deal with a succession of political and economic upheavals beyond their borders. The fall of the Berlin Wall has created undreamt-of opportunities in Eastern Europe and beyond. The opening of India and China to foreign companies has caused a lot of excited slavering over the prospect of billions of new consumers.



The wild East

Sources: IMF; World Bank; OECD

Even during the recession of the early 1990s, foreign direct investment (FDI) in the developing world continued to boom. In 1993, for example, developing countries attracted a record $70 billion in FDI, nearly twice the amount they had received in 1991 and almost as much as the world's total FDI in 1986, according to the United Nations study group on transnational corporations. The rush to China has been particularly dramatic. China is now the largest recipient of FDI in the developing world, and the second largest in the world after the United States. Last year its economy as a whole grew by an estimated 11%, and its "special economic zones" in the south and along the coast much more quickly than that.

In this booming, bustling but worryingly unpredictable part of the world, the word on the lips of most multinational managers is "localisation". Locals are much more likely than foreigners to have the personal connections (known in China as "guanxi") so essential to doing business in the region. They also have the sort of insider's knowledge which saves them from making crass mistakes. The trouble with "localisation" is that good locals are difficult to come by. The Hong Kong labour market is one of the tightest in the world. On China's mainland, the dearth of managerial talent and technical skills that seems to be a common feature of communist and ex-communist countries was exacerbated by the cultural revolution, which thinned the ranks of people between 35 and 45.

Despite these problems, some companies are adopting a policy of "Chinese at any price", scouring the Chinese communities of the United States and Asia for suitable ethnic Chinese. The talk in Hong Kong is full of ABCs (American-born Chinese) and "returnees" (people who fled the People's Republic but are now returning as company men.) The region is seeing the birth of a new sort of expatriate: an ethnic Chinese who has been educated in western business schools and spends his time moving from one lucrative posting to another. According to a local joke, in Hong Kong you hire the company, not the other way round.

Nevertheless, locals do not always go down well with the people they are supposed to impress. Mainlanders do not necessarily take to these "ABCs with their MBAs", particularly when they have little grasp of the local dialect. They like to negotiate with "grey hairs" rather than with people in their early 20s who still reek of Harvard Yard. Companies find that to clinch a serious deal with the Chinese authorities, they have to wheel out heavyweight senior company figures.

Even if they can find the right locals, multinationals have to remember that they are worldwide companies which happen to be in China, rather than Chinese companies which happen to have a global arm. They have to make sure that what they transplant into Chinese soil is the way their company actually does things, not some bastardised version of it; and they have to make sure that they do not get involved in any practices that might damage their brands. Headlines about corruption or pollution can hurt the company's image the world over. Involvement in shady deals can also land American managers in jail. In other words, multinationals in places such as China need powerful control mechanisms; and that probably means expatriate managers.

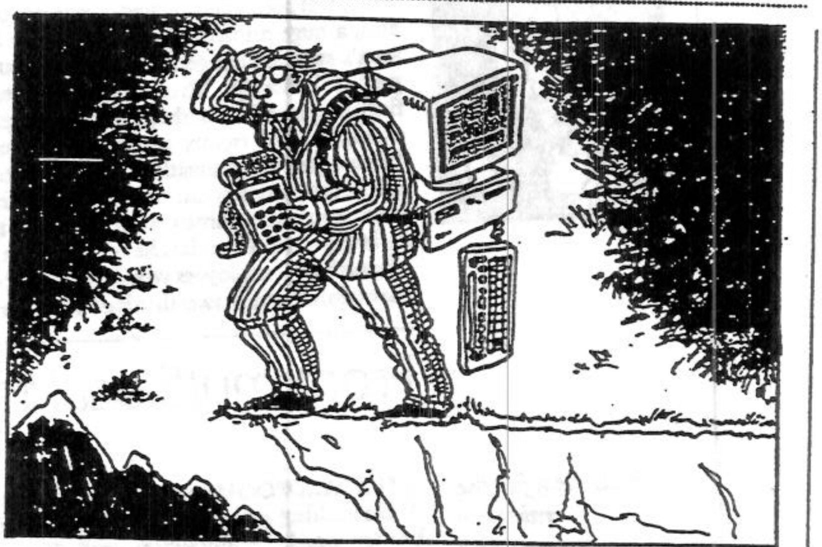
The expat rides again

Price Waterhouse, a consultancy, calculates that some 450,000 expatriate managers are now working in China, and predicts that their number will continue to rise for the rest of the decade. Expats are expensive to employ, requiring generous salaries, extensive insurance and four or five spells of paid leave away from China every year. Critics say they spend the first two years of their posting getting up to speed and the last year packing up, leaving only a brief period when they are on top form. Sensible firms take a lot of trouble to ensure that potential recruits are well-informed about the hardships of a China posting, and that both they and their spouses are up to it. The managers, for their part, may be far from keen to go: in these days of downsizing, delayering and short-term contracts, many managers no longer see a foreign posting as a safe step on their career ladder, and many of them have spouses whose own career would be interrupted by a move.

The "expats" come from a rather broader range of backgrounds than the traditional members of the species. Some read Chinese at university; some are specialists in some of the many areas in which China suffers from skill shortages, such as accountancy and engineering; some are middle-aged managers keen to become entrepreneurs. They need to be senior enough to impress an age- and status-conscious society. They also need to exercise sufficient clout with head office to ensure it does not expect too much too soon, and to get the resources they need. In turn, head office needs to give them plenty of leeway to avoid compromising their local credibility. AT&T and Owens-Corning Fibreglass are going one step further and setting up stand-alone units to go after business in the region.

The expat managers need to be willing to work themselves out of a job. The art of management in this region is very much the art of transferring knowledge, implanting western business methods in Chinese minds. PepsiCo plans to establish a training centre in Shanghai for 50 people a year; it already sends high-flyers out of China for up to two years' training. Price Waterhouse's training centre in Shanghai, bristling with the latest multimedia equipment, can take only one applicant in ten.

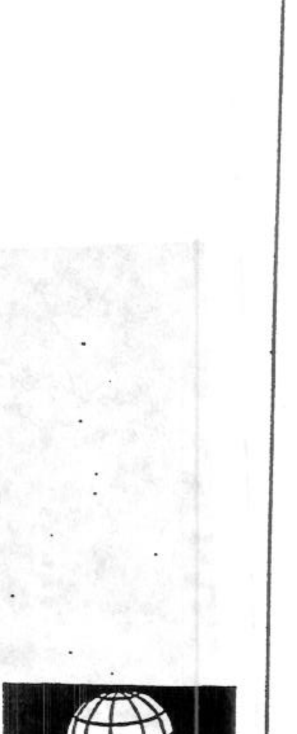
Trained employees are so valuable that companies have devised various forms of "golden handcuffs" to make sure that they do not defect to competitors. One popular method is to offer them plenty of travel abroad. Another is an arrangement with a local housing co-operative which leads to ownership of their house if they stay loyal for a decade or so. Companies are also having to accommodate the Chinese taste for a predictable career ladder and an elaborate hierarchy. Even as they are flattening their hierarchies in the West, they are elongating them in the East.



Another—and until recently almost the only—way to become an insider is to form a joint venture with a local business. Managing these arrangements is tricky. It requires heavy investment in capital and training, and leaves companies prey to local malpractice. As a short-cut, some multinationals are forming alliances with blue-chip companies which have a long history of operating in the region, such as Jardine Matheson and the Swire Group. Others have been throwing themselves into the arms of the large Overseas Chinese family firms which operate throughout the region. Wal-Mart, for example, has an alliance with Thailand's Charoen Pokphand (CP) Group, Mitsubishi with Indonesia's Lippo Group and Pearson with the veteran Hong Kong television tycoon, Sir Run Run Shaw.

Most successful companies have observed two all-important rules in forming joint ventures and alliances. The first is to get in early. One reason why Coca-Cola has 16% of China's soft-drinks market (compared with Pepsi's 6%) is that it moved quickly to tie up the best local firms in the most promising markets. The second rule is to manage risk in such a way that the bulk of it falls on your partners. Again, Coca-Cola has played this game masterfully. If the local bottler fails to come up to scratch, the company simply takes its formula and brand name somewhere else.

But managing in China can be intensely competitive—and not only because of other western multinationals. Gaze out over Hong Kong at night, and among the bright neon signs advertising names such as Citibank and Pepsi familiar to every westerner you will see equally bright signs carrying less familiar names such as Lippo and CP. These emerging Asian multinationals have deep pockets, burgeoning business interests and cosy relations



with local politicians. Below these large conglomerates, a huge number of smaller family firms (Taiwan's stock exchange alone lists 400,000) is competing furiously. These firms are remarkably flexible, providing perhaps the most sophisticated examples of a trendy management concept, the "networked" organisation. If they need global reach, the patriarch can call on his allies in the Chinese community throughout the region; but if agility is the order of the day, he will delegate the job to a handful of employees who happen to be close to the action. Perhaps two-thirds of the investment on the mainland comes not from multinationals of any kind but from these smaller Chinese firms.

To have any chance of prospering in such competitive environment, foreign multinationals will have to become part of the local business scene, promoting local people to senior positions, forming close relationships with local patriarchs and learning to think like a local Chinese business—as well as remaining true to themselves. The next article looks at how such companies ensure that the many cultures within the organisation do not clash but mix to best effect.

From multilocal to multicultural

This way for the rainbow corporation

IN BOARDROOMS around the globe, managers are mulling over a new concept summed up by a beastly word: "multiculturalism". The idea, put rather more briefly than most of those managers would, is that the world's centres of economic activity are becoming more dispersed, and that multinationals should therefore put a wider variety of people on to their boards to broaden their perspective and alert them to changes in demand and fashion. Such a policy might also improve their chances of recruiting and motivating able people from the many countries in which they operate. Bright young Chinese or Indians, for example, might give the cold shoulder to firms that seem to be reserving all their top positions for whites.

Many multinationals like to boast that they are now choosing their senior managers from around the world. L'Oréal, a French cosmetics company, is run by a Welshman, Lindsay Owen-Jones; Heinz, an American food company, by an Irishman, Tony O'Reilly; and Ford by a Scotsman, Alex Trotman. Schering, a thoroughly German chemical company, is headed by an Italian, Giuseppe Vita; McKinsey, an unmistakably American consultancy, by an Indian, Rajat Gupta. The boss of Nestlé, a Swiss food company, is a German, Helmut Maucher, and fewer than half its board members are Swiss. Shareholders, too, increasingly hail from around the world. Over the past decade the number of countries where the world's top ten multinationals' stock is traded has risen from 58 to 70.

Both European and American companies claim to have an advantage in the multicultural stakes. The Americans recall that their country is the melting pot of nations, and that they can do business in many developing countries without triggering unpleasant memories of colonialism. The Europeans point to their long and distinguished history of multicultural management, citing examples such as Royal Dutch Shell and Unilever. Because many European countries are small and closely-packed, their citizens have had to become linguists—which the Americans have conspicuously failed to do—and to develop an awareness of different cultures.

Much of this is bunk. Beneath the surface, many multinationals remain stubbornly monocultural. The proportion of foreign-born board members of America's 500 leading companies in 1991 was 2.1%, the same as ten years earlier, according to Susan Schneider, a professor at the Insead business school near Paris. In Switzerland and Israel, top managers get to know each other through compulsory military service. In France, half the heads of the top 200 companies were educated in the same six *grandes écoles*; only a handful have any experience of managing abroad. And even the most conspicuously global companies are suspected of national favouritism. European staff know Ford Assembly Operations as "For Americans Only".

Multiculturalism can also be overdone. A firm's internal culture may be its most valuable resource. Visiting one of the world's great companies is rather like visiting a cult: company people have their own local heroes (such as the founder or the current chairman), their own local war stories (about how they beat the competition) and their own language (Ford workers seem to have nicknames for everything at their Dearborn headquarters). Companies need to make sure that, in adopting multiculturalism, they do not dilute their own culture.

By their nature, multinational companies are part of the international business world. Managers read international (which usually means American) business books and attend international conferences. Many blue-chip companies, including Thomson, Alcatel, L'Oréal, Rhône-Poulenc, Continental and ABB, have adopted English as their official language and use it even in their home countries. Companies habitually put high-flyers into a series of foreign postings to broaden their experience and give them an international perspective. Once a recruit from a developing country has moved through half a dozen different postings abroad, his memories of his country of origin may have become faded and out-of-date.

Still, many companies, particularly in banking and chemicals, are forced to take multiculturalism seriously because of mergers and acquisitions. Deutsche Bank, for example, is keen to gain cultural as well as strategic advantage from its takeover of Britain's Morgan Grenfell. The continued integration of the European market will no doubt lead to many more cross-border mergers. Percy Barnevik predicts that by the end of the decade 30% of big European companies will have to close or merge.

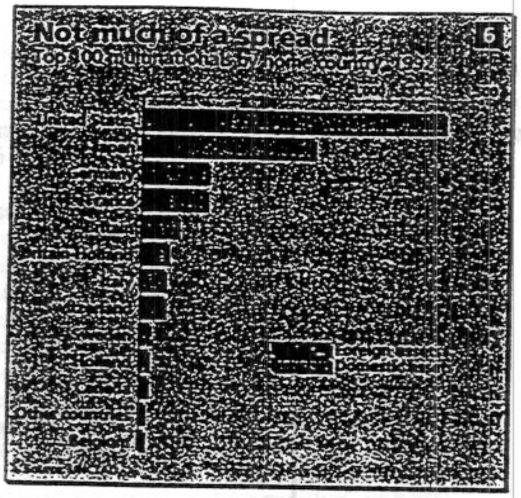
Some of America's leading companies, too, have become aware of the need to create a more multicultural workforce. Corning, a technology company based in a small, predominantly white town in upstate New York, has made formidable efforts to attract and retain more black workers, introducing "diversity awareness" programmes in local schools and even getting the town wired for a black television channel. Avon, a cosmetics company which is doing particularly well in emerging markets such as South America and China, has set up a "multicultural participation council" to encourage diversity among employees. Digital Equipment, a computer company, sponsors events such as Hispanic Culture Week.

Companies have also been putting pressure on American business schools to become more international. That means recruiting staff and students from as many countries as possible, and offering courses that will help shake off the image of the "ugly American", unable to speak any foreign languages and utterly insensitive to cultural differences. Recruiters doing the rounds of business schools are particularly keen on hiring non-Americans who combine the advantage of an American business school education with an insider's knowledge of another culture.

All this, and political correctness too

Western multinationals are also trying to polish up their image in another way: by embracing lily-white business ethics. They want the world to forget the days when the price of doing business in some countries was cosying up to a government of doubtful moral standing. Most well-known companies have high-minded internal mission statements and codes of conduct; they also sign up to international agreements such as the International Chamber of Commerce's Business Charter for Sustainable Development. A few even have full-time ethics officers who monitor the company's conduct.

It is tempting to dismiss all this as a public-relations exercise, but such codes of conduct do sometimes affect the way a company runs its day-to-day business. Levi-Strauss, an American clothing company, cancels contracts with suppliers if it finds they employ child labour or overwork their staff.



Western chemical companies, realising in the late 1980s that pollution scandals could put them out of business, became extremely vigilant in policing the industry. Computer companies such as Digital, Compaq and IBM have pushed for higher environmental standards. Accountancy firms such as Price Waterhouse are helping post-communist countries set up modern accounting systems. Oil companies such as British Petroleum and Arco guarantee that they will build schools and airports and act as green watchdogs, in return for permission to drill for oil in places such as Siberia and Alaska. It seems that behaving like good corporate citizens makes eminent business sense.

Another benefit that multinationals confer on their host countries is that they tend to impose international standards wherever they go—or have them imposed on them by global shareholders and media. On the whole they find it easier to operate one set of rules everywhere around the world, rather than to apply different rules in different places. So multinationals clamour for more global—and usually higher—standards, partly because it makes their lives easier, partly because it imposes the same standards on their competitors.

This strategy seems to be paying dividends, at least in terms of public relations. Multinational companies seem to have shaken off their once-grubby, corrupt and exploitative image. The United Nations, which used to try to control multinationals, now seems to regard them as conduits of good practice and agents of international integration. A recent report from the UN's committee on transnational companies pointed out that they are now shifting increasingly sophisticated operations to developing countries: 3M, for example, is producing tapes, chemicals and electrical parts in India, and Hewlett-Packard is designing computer memory boards there. The multinationals' apologists also argue that, even when such firms move to developing countries to cut their labour costs, they invest heavily and introduce superior management and better technology, thus helping to upgrade the host economy and deepen global integration.

But not every company is equally comfortable in this brave new multicultural world. And one group is finding it harder than most to make the transition: Japan's multinationals.

Japan looks west

Time for a little corporate role reversal

IT DOES not seem long ago that western businessmen could not hear enough about Japanese management, with its "lean manufacturing", "just-in-time production" and "continuous improvement" (*kaizen*). Less than 20 years later, the Japanese are returning the compliment. They want to know how western companies are dealing with "middle-management spread"; how they are using new technology to process information and improve communications; how they motivate people in far-flung factories; and how they weld those factories together into worldwide networks. Sophisticated Japanese managers are now littering their conversation with words like "downsizing", "delayering" and—a particularly difficult one for them—"re-engineering".

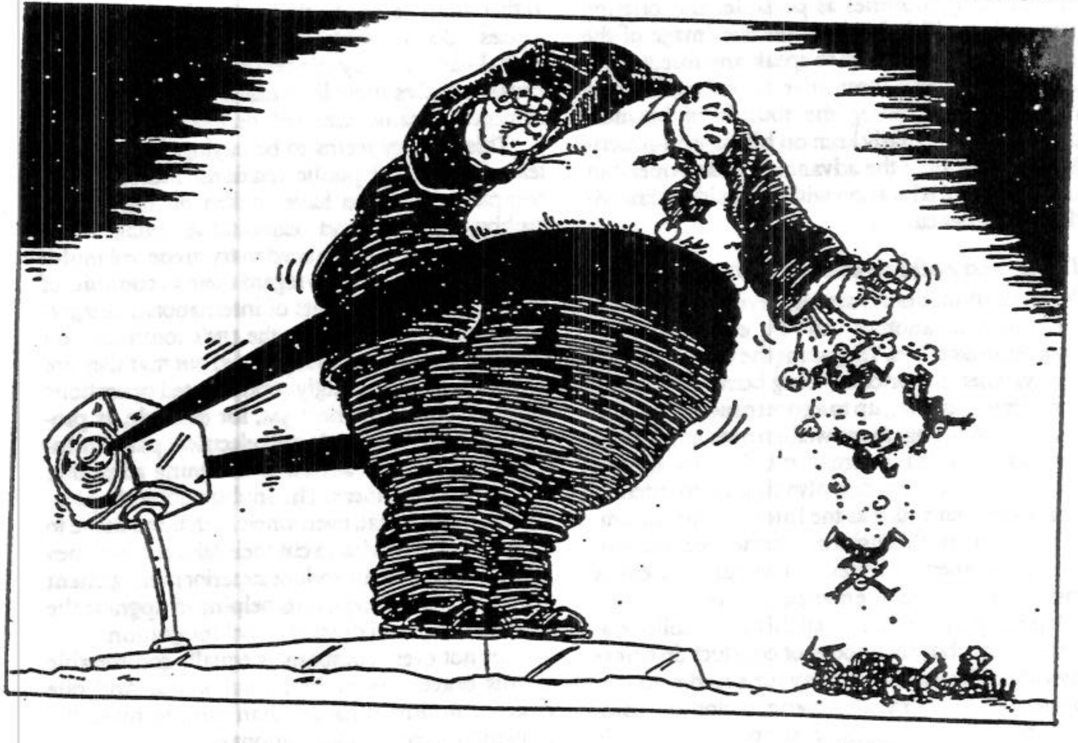
Japanese multinationals are right to look west. In 1992 a study published by the Japan Productivity Centre, a semi-offical body, suggested that Japanese companies will have to sack 39% of their staff to match the productivity of American firms. In particular, Japan's middle managers are rapidly becoming a burden rather than a strength. Whereas they used to be cheap and flexible, they have become expensive, thanks to seniority-based pay, and rigid, thanks to their computer-phobia. Western companies contract out information-processing to sleekly managed specialists; Japanese ones still employ rows of deskworkers, some of them with abacuses. Yet, until recently, the ranks of white-collar workers have been growing. Between 1984 and 1992 the proportion of administrative and sales staff at 50 companies studied by Bain, a consultancy, increased from 29.2% to 33.5% of the workforce. Japan's companies may need to master the art of "lean management" if they want to survive.

Companies which were once content to produce everything at home are also having to learn how to manage their increasingly far-flung operations. Fear of exclusion from the world's most lucrative markets has persuaded many of them to set up factories in the United States and Europe; and increasingly stringent rules in those countries on "local content", combined with accusations that Japanese factories abroad are little more than "screwdriver plants", have prompted them to add some research and development.

Although Japan's overseas investment has slowed dramatically in the past few years, the strength of the yen, which has risen by 20% against the dollar since the start of the year, is now forcing even the most stick-in-the-mud companies to consider shifting production abroad. In industries such as electronics, the market also increasingly calls for short runs of carefully targeted products rather than standardised global wares, forcing the Japanese to move their research and development facilities nearer to foreign markets. The question is whether Japanese multinationals can transform themselves from top-heavy exporters into nimble global networks without sacrificing the very qualities that made them successful in the first place.

Slim is beautiful

Even the most successful Japanese companies are desperately trying to slim. Toshiba, an electronics giant, is reducing its headquarters staff by 30% over the five years to 1997, delegating more power to front-line workers. The latest war-cry in its headquarters is "small, simple, speedy, strategic". Fujitsu claims that it has cut its workforce at headquarters by a fifth since 1992. Nissan is aiming for a



reduction of 5,000 in its staff by the end of this year, with most of the job losses falling on white-collar workers. Even Canon, which is busy concentrating all its administration on its gigantic corporate headquarters on the outskirts of Tokyo, has long-term plans for a much slimmer operation.

Japan's multinationals are also trying to find ways to square their traditional paternalism with the efficiency needed to stay competitive. They are becoming much more ruthless about identifying low-achieving middle managers and moving them to sales or even production jobs, or shifting them into other parts of the "*keiretsu*", or family of companies. Above all, they are redefining what they mean by "life-time employment", making it clear that the system only ever applied to part of the labour force, and whittling down that proportion as far as they can. Japanese banks have introduced a system of "up or out": those who do not make the grade by the age of 40 are sent away to run local branches. Toyota has taken to moving 50-year-old managers from supervisory jobs to "individual work". The more desperate companies are introducing "voluntary" early retirement.

But this is still a far cry from rationalisation as a western company would define the term. Companies still prefer to put a freeze on new recruitment rather than weed out middle managers. Nissan, which used to hire 1,500-2,000 new recruits a year, plans to recruit only 55 newcomers this year. Graduates anxiously study companies' age profiles, which are published annually. Women are eased out as soon as they get married, no matter how good they are at their jobs.

The management of many Japanese multinationals has barely begun to come to grips with information technology. Although the Japanese pioneered the use of the fax, they are much less happy with computers, voicemail and video-conferencing. This is not just because the Japanese language does not take happily to the keyboard: there are strong cultural reasons as well.

The Japanese invest a great deal of effort in developing personal networks and attach particular importance to face-to-face contact: this is, after all, a country which boasts door-to-door car salesmen. Some Japanese firms are so worried about the potential impact of e-mail on the company hierarchy that they are experimenting with it in their foreign subsidiaries first, according to Douglas Shintano, head of EDS's Japanese office.

Go global, young man

The big Japanese multinationals accept that they must learn the art of global management. Toyota prides itself on having started to convene shareholder meetings outside Japan. Nissan is keen on creating a "global team spirit". Canon is trying to produce a "global mind set" among its employees, and gives them free videos showing what the company is doing all round the world. Toshiba is reorganising every part of its operation on a worldwide basis, partly to get rid of duplication, partly because, it argues, there is no longer such a thing as a purely Japanese business problem. The Japanese are also learning to manage their overseas operations with a much lighter touch. Canon's 2,000 employees in the United States include only 20 Japanese, and half the board members of Canon UK and Canon USA are non-Japanese. Worldwide, Canon now employs almost as many foreigners as it does Japanese.

GLOBAL
VISION
SPECTACLES

The best companies are also moving research and design facilities abroad. Matsushita has a regional training centre in Singapore and does research and development in air-conditioning in Malaysia. Canon and Toshiba have both produced breakthroughs in audio technology by setting up laboratories in Britain. Yoshikazu Kawana, Nissan's executive vice-president, says that his firm wants to use the entire world as a "knowledge base". Hajimi Mitarai, Canon's president, says that the company has to learn from best practice in each and every one of the countries where it operates. Mr Mitarai has blazed a trail by getting degrees from both MIT and Stanford.

But many Japanese multinationals still find it difficult to manage foreigners, particularly above the level of production workers. The main problem lies with exporting their consensus-based approach. This has served them well at home, institutionalising perpetual discussion, obliging managers to gain the enthusiastic support of their workers and encouraging all employees to collaborate across functional boundaries. But hitherto the two elements of this approach, "*nemawashi*" (consensus-building) and "*ringi*" (shared decision-making) have depended on close physical proximity and shared cultural values. Such a package does not travel well.

What the Japanese have succeeded in exporting is their unique approach to manufacturing. In places as far apart as Wales and Tennessee, local workers are using "Japanese" methods to produce Toyotas and Nissans for their domestic markets. Many of them regard the Japanese as much better employers than their western counterparts. These workers have also been adapting Japanese methods to local circumstances. Fujio Cho, a managing director at Toyota in Japan, claims that the great virtue of "lean production" is that it is not just self-correcting but self-adapting. The whole point of the system is to expose the problems below the surface, and then let the workers solve them in their own way. Both the problems and the favoured solutions

differ strikingly from country to country. American workers prefer technological fixes, whereas Japanese ones tend to work harder or redesign their working methods.

The idea seems to be that Japanese managers should set things rolling and then withdraw gracefully, leaving the field to the locals. But at that point the model stops working as intended. Japanese multinationals' foreign subsidiaries depend much more heavily on expatriate managers than do their western counterparts. They also promote far fewer foreigners to senior positions and involve them in fewer important decisions. Such criticisms suggest that Japanese multinationals have a serious problem. The "rice paper ceiling" on which foreigners hit their heads deters many able people from joining Japanese firms, and demotivates existing staff. It also prevents Japanese companies from making full use of their employees' abilities and ideas.

For all that, though, western multinationals underestimate the Japanese at their peril. Japanese companies excluded from the cosy business club which runs Japan have done better when moving production abroad. Sony's famous boss, Akio Morita, even moved to the United States for a time so that he could better understand his company's biggest market. The Japanese have also learned some lessons from their failures in Hollywood. And Japan's trading companies, with their international networks and information banks, have always been a huge asset to Japanese global ventures, particularly in Asia.

Besides, as Nissan's Mr Kawana points out, Japan is changing. Many younger Japanese people are no longer willing to tie themselves to one company for life. They are also much more curious about the outside world. Thirty years ago few Japanese had spent much time abroad; now anybody who wants a senior position has to gain experience overseas. But are things changing fast enough to ensure that Japanese companies will succeed at becoming truly global? The next article will examine the formidable array of skills required to achieve this most difficult of multinational tricks. ■

Only communicate

Thoroughly modern corporations know no borders

MULTINATIONAL managers might be forgiven for complaining that they are being given ever more to do, and ever fewer tools with which to do it. Where once they ran local fiefdoms, they are now co-ordinating global networks. At the same time they are losing many of the control mechanisms which they once considered essential to their job. The need to reduce overheads has left fewer people to do the managing, and the bigger role for alliances and joint ventures has removed many functions from formal managerial control.

Most multinationals are dealing with this problem by shifting from formal to informal control mechanisms. Formal methods are costly and time-consuming, and tend to produce a less than enthusiastic response. It is far more effective to get people to believe in the company's values than to keep issuing them with instructions and keeping a close eye on their performance. But even as they dismantle their more formal controls, some companies are weakening their informal ones as well. Pressure of time and money is forcing them to compress training courses which play a useful role in creating a team spirit and building up corporate networks. Global expansion—some of it through mergers and acquisitions—combined with the need to localise management means that managers no longer necessarily share the same set of basic assumptions. More problematic still, many managers are, in effect, being asked to help make themselves redundant. Country bosses are having to transform themselves from local barons into loyal lieutenants, with a consequent loss of power, and middle managers are required to develop systems which will make middle management unnecessary.

Moreover, informal control mechanisms do not always help managers to resolve the tension between the local and the global which is at the heart of all multinational management. The method recently chosen by a number of companies to deal with that is surprisingly old-fashioned: matrix management. This is an elaborate and bureaucratic system of formal controls which means managers report not to one but to two bosses, a global (or regional) one who is responsible for a product or function, and a geographical one. The matrix is supposed to break down regional and functional divisions; to force people to think in more than one dimension; and to provide decision-makers in the organisation with a richer flow of information than they would get under a simpler system.

The renaissance of matrix management owes much to the advocacy of ABB's Mr Barnevik, who seems to have succeeded in operating such a system without dulling the imagination or restricting the initiative of his front-line workers. But even in Mr Barnevik's skilful hands the system has not been an unmixed blessing, and other companies have found that the drawbacks outweigh the advantages. Robert Palmer, boss of Digital Equipment, blames the matrix structure for severely delaying his company's shift from minicomputers to personal computers. Many other senior managers agree that the system is time-consuming and bureaucratic, that it institutionalises dithering, and that it leaves everybody disappointed, since nobody can satisfactorily serve two masters.

Companies are getting more joy out of humbler techniques, such as training programmes, internal conferences and annual jamborees. Instead of parachuting expats into remote parts of the world, many companies are instead putting locals through intensive training courses before returning them home to run the show there. AT&T brings foreign-born managers to the United States for two-year stints. Gillette, a Boston-based maker of personal care products, runs local courses in the 28 countries where it has offices, then sends the trainees to Boston, London or Singapore for 18 months. At McDonald's Hamburger University in Oak Brook, Illi-

nois, up to half the 250 students are from outside the United States, and simultaneous translation is available in 22 languages.

Successful firms such as Nestlé and ABB have developed a networking culture in which employees from different parts of the organisation are constantly being thrown together. Most companies feel they get their money's worth from seemingly extravagant jamborees for their managers from all over the world. Such face-to-face meetings help to form personal networks among people who usually just talk to each other on the phone, and instil corporate loyalties among managers who might otherwise be inclined to defend their own little patch. They also help to reveal common problems. Ford says that one of the things which persuaded it to become truly global instead of remaining multi-domestic was that at the company's training college so many senior managers from different parts of the world kept coming up with the same complaints.

Multinationals now spend much more time and effort trying to explain to their employees what they are doing, and winning their support. One widely used device is to craft corporate vision and mission statements, often drawing in large numbers of staff. The vision may be fluffy and the mission anodyne, but the process of creating these statements seems to reinforce corporate ties. There are many other ways, too, of involving the workforce. Ford, for example, has an internal television network which keeps employees informed about the company's business, giving the stock prices of the world's car companies and the latest motor-racing results. Ford employees can also log into a central computer network and find out about corporate policies, new models and a thousand other things. Philips celebrated its 100th anniversary in 1991 with a televised worldwide conference at which the chairman, Jan Timmer, explained the company's difficulties and prospects to the entire workforce. Other examples abound.

Managers spend a huge amount of time communicating with each other as well as with their subordinates. William Bruns, a professor at Harvard Business School who has recently studied managers from 55 multinationals for a forthcoming book on corporate control, is amazed at the amount of communication that goes on, particularly over the telephone. Corporate men and women consult each other almost obsessively about people and problems. In addition to telephones and faxes, many of them now also have voicemail and electronic mail to keep in constant touch with colleagues around the globe. Groupware, Lotus Notes, electronic whiteboards and other software programmes provide further means of working together at a distance, and videoconferencing allows face-to-face contact with people a continent away. Videoconferencing technology, once cumbersome, is developing rapidly and becoming increasingly popular. Managers at Daimler-Benz's new factory in Alabama hold a two-hour teleconference with head office in Stuttgart every day.

But all this new technology is not doing away with the need to trot the globe; if anything, it seems to have increased it by putting ever greater emphasis on communications. In his study of multinational managers' habits, Professor Bruns was as surprised by the amount of travelling as by the amount of talking they do. Many of them spend three out of every four weeks on business trips. Some have their timetables worked out 12 to 15 months in advance, and elaborately co-ordinated with those of their close colleagues.

Do managers really need to travel this much? They insist they do, arguing that it helps them to form closer personal ties with their colleagues abroad and to understand their behaviour better, particularly if they come from a different culture. Dinners and late-night drinks thus have a prominent place in their elaborate timetables. Managers say they have to meet people in person when deals are being negotiated, new staff are being hired or controversial decisions pushed through. Above all, they feel that travel is essentially about trust: they need to reassure themselves that the bonds of trust between head office and the rest of the company remain in place.

Multinational managers underestimate the importance of trust at their peril. Chan Kim, a professor at the Insead business school, and his colleague Renée Mauborgne conducted a survey of 3,500 companies and came up with the finding that, to be successful, managers need to put strong emphasis on "procedural justice". Multinational management is a matter of making continuous trade-offs between the global and the local. These need to be seen to be for the greater good of the company, not for the benefit of any particular group. If decisions are properly explained, subordinates will acquiesce in them; if not, they will subvert them.

Despite teething troubles, multinational managers are clearly learning how to run the flat, nimble, globally-integrated organisations which are replacing yesterday's cumbersome and bureaucratic giants. But why bother? Why not simply take the revolt against gigantism to its logical conclusion and break up the multinationals into their component parts? Why not rely on alliances for global reach rather than incur the trouble and expense of headquarters and global managers?

Alive and kicking

The death of the multinational has been much exaggerated

RUNNING a multinational is a difficult job, and getting harder. Not only are big companies getting ever more complex, but boundaries between them are rapidly dissolving. Many firms are forging much closer relations with their suppliers, giving them sensitive information and helping them to run their operations more effectively. At the same time companies everywhere are forming alliances: computer firms with telecommunications-providers, airlines with hotel chains. This means that businessmen now have to think about groups of firms rather than just their own.

They also have to get used to living with ambiguity. Today's competitor may become tomorrow's ally, or vice versa; indeed a company may be collaborating and competing with another at the same time. Jordan Lewis, a consultant who specialises in alliances, points out that the skills required to manage them are becoming mainstream, not only because alliances are now common but also because, even on the inside, firms are beginning to resemble networks of alliances.

Trust me

The second reason why multinationals are becoming harder to run has to do with trust. Because of the widespread shift to informal controls, managers and their staff need to trust each other. Because of the importance to corporate success of ideas and knowledge, everybody must be willing to share the fruits of their thinking. Because of closer relations with suppliers and other companies, businesses must be prepared to exchange trade secrets and work together without formal contracts. Yet at the same time the pressure of competition is undermining trust. Alliances can be severed. Jobs can disappear in a wave of downsizing and delayering. Time is often too short to develop long-term trusting relationships. In a multicultural environment, people can no longer rely on shared backgrounds and tacit understanding.

The people who run these organisations must constantly perform near-miracles. Most of them are extremely well-rewarded, sometimes to the point of arousing public indignation. But their huge salaries increasingly reflect huge workloads. Michael Miles, a former head of Philip Morris, a food and tobacco company, used to get into the office at six o'clock every morning; asked what he did besides work, he replied "sleep". Percy Barnevik manages on four hours' sleep a night and travels for up to 200 days a year. He says he is normally in his office on two days a week: Saturday and Sunday.

So why, if the work is so hard, the job so difficult and the world so critical, do these bosses carry on? If you asked them, you might get a different answer from every one of them. But all of these answers would probably have one thing in common: a strong underlying belief that, for all their troubles, multinationals are here to stay. This survey will end by examining the evidence for that belief, and conclude that it is well-founded. Economies of scale may no longer be as important as they were, but another competitive advantage of multinationals has become crucial: global reach.

At the beginning of this year Ford formally bet its future on the "global car". General Electric and Gillette take an all-or-nothing stance: either they are the world's number one or two in any particular product line, or they sell it. Firms such as Unilever or Nestlé, which are famous for their sensitivity to local differences, are trying to have the best of both worlds by adapting global products to local tastes but also establishing product groups with global responsibilities.

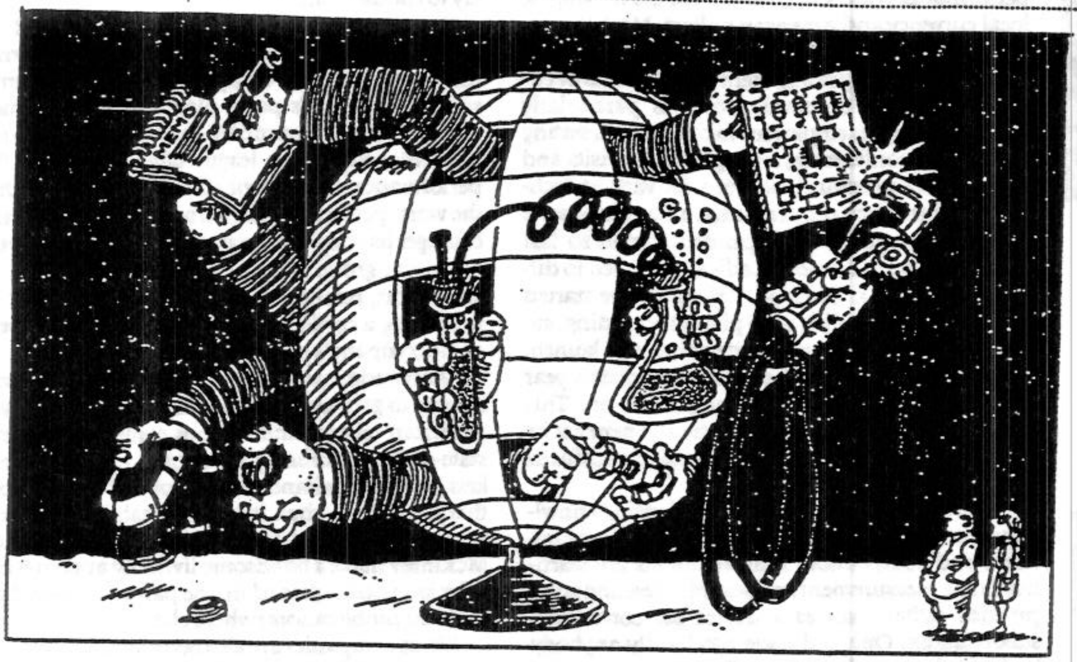
In some industries, such as jet aircraft and semiconductors, companies have always needed to sell globally to recoup their investments. The number of such industries may be increasing as the cost of innovation soars, for example in pharmaceuticals and telecommunications. In the 1960s, a company such as Siemens had to do no more than capture half the German market to recoup the cost of developing an electromechanical switch; soon it will have to capture 20% of the global market to make such an investment worthwhile. But even in industries requiring much less heavy investment, global companies have a formidable advantage over their local rivals—provided they are flexible enough to adapt their global products to local circumstances.

Properly integrated multinationals need only one team of people to design and develop a product for the entire world market. Ford's designers around the globe are now connected by interactive computers which enable them to manipulate the same image and talk to each other by videolink. Philips developed its wide-screen television using a team some of whose members were based in Eindhoven, Holland, and some in Singapore.

Multinationals can also use the same basic advertising pitch to sell the product around the globe, refining the idea in the light of worldwide experience, adjusting it to tricky local markets and all the time implanting the brand's name ever more deeply in the mind of consumers. OC&C, a London-based consultancy, has calculated that the advertising and promotion costs needed to persuade a customer to try out a new product are, on average, 36% less for a "recognised" brand than for an unfamiliar one. Finally, they can use their international presence to roll out the product worldwide.

The world is our oyster

Many national companies feel that they have no choice but to become international because that is the only way they will be able to defend their home market. Ford was forced to introduce its "global car" because it found that a collection of semi-autonomous national companies, duplicating everything from research to advertising, could not compete with more centralised Japanese companies. Global companies can also deploy their resources more effectively than national ones. Gillette prides itself on switching production from factory to factory around the world to cope with fluctuations in demand, thus keeping its spare ca-



pacity to a minimum.

Global reach can save time as well as money. It is now common practice for law firms in New York to fax work to their partners in Hong Kong at the end of each day, getting the results back the next morning. Texas Instruments, an American computer company, has established teams in Bangalore in India as well as in Texas, so that it can work on a problem 24 hours a day.

Conversely, global reach can stretch out time to the multinational's advantage. Companies such as Gillette prolong the life of a product by launching it in the fashion-conscious rich world first and in less affluent and developing countries at a later stage. The cycle is getting shorter, because emerging companies are now so good at producing copies, but the strategy remains the same. Multinationals can also extend the life of their equipment by moving it on to the less developed parts of the world. Gillette claims that some of its equipment is on its fourth or even fifth time round.

Risk, too, is easier to deal with for a global company. Not all products succeed, but large R&D budgets can absorb failures more easily than small ones. Products that sink in one part of the world may swim in another. Companies can experiment with changes in their production processes in markets which are large enough to matter but not so large that they risk bankrupting the company. Gillette is currently introducing just-in-time production methods, linking production much more closely to sales, in Scandinavia.

Global knowledge machines

Moreover, multinational companies are usually better than their more parochial rivals at the all-important business of creating, gathering and cross-fertilising knowledge. They enjoy access to a larger pool of management talent, a wider range of skills and a greater variety of perspectives; and they are likely to know more about such factors as consumer needs, technological changes and competitors' moves. They can get access to knowledge from all over the world, particularly of the sort that requires an inside view. Philips is moving more and more of its electronics business to South-East Asia, arguably the world's most creative and competitive electronics market. Ford is making small and medium-sized cars for the world market in Europe, because that is where the demand for them is concentrated—and so is the expertise in making them. Hoechst, a German chemical company, has shifted the bulk of its genetic research to the United States.

Multinationals' knowledge advantage extends to being able to respond to changes in taste around the world, a particularly valuable skill now that standards are being set in places such as Hong Kong and Tokyo as well as in Paris and New York. Philips decided to produce its new up-market large-screen television when head office received a memo from Jack Lau, head of its marketing team in Hong Kong, pointing to the dramatic increase in demand for such products in the region.

A further strength is the ability to combine skills and knowledge from different parts of the world. At its simplest, this can mean putting American brains together with Filipino hands, as Nike does so successfully. More elaborately, lift companies look to Europe for their drive technology and to America for their advanced control systems and software. Electrolux has a research laboratory in Finland, a development centre in Sweden and a design group in Italy. The more adventurous multinationals are turning themselves into knowledge-brokers, bringing together people who might not normally meet: Italian designers and Japanese computer experts, perhaps, or Taiwanese advertisers and German chemists. Some companies are trying to formalise this skill by appointing "chief knowledge officers". But most simply get on with it.

The best companies aid the process of moving skills and knowledge across borders by making sure that all their workers can get instant access to company-wide information. At ABB, for example, employees can tap into an information system, Abacus, which collects performance data on each of the

This survey has received help from many quarters, but has benefited particularly from conversations with Christopher Bartlett, William Bruns, Yves Doz, John Stopford and Chan Kim. Among published sources, it has drawn especially heavily on the annual World Investment Reports of the Transnational Corporation and Management Division of the United Nations; on working papers of Insead's corporate renewal initiative; and on various case studies prepared by Harvard Business School, Wharton Business School and Insead.

company's 4,500 profit centres and compares their performance with budgets and forecasts in both the local currency and American dollars. Most companies also move their employees around fairly frequently, either by posting them to new places or encouraging them to travel. Nestlé is particularly good at getting its employees to collaborate, setting up internal conferences, arranging short visits and rotating key personnel between its various technical and research centres. IKEA, a Swedish-based furniture company, has expanded abroad so fast that it has had to develop a radical approach to diffusing in-house knowledge. New stores are started by a team of specialists who plan the building, supervise its construction, mastermind the launch, run the store and train the local team; after a year they hand over to the locals and move on. This means that the elite team gets better and better at its job, while IKEA's core skills are quickly diffused around the world.

Two other techniques for disseminating knowledge are proving particularly effective: measurement and competition. Multinationals are learning to use measurement as a way of spreading good practice rather than as a device for controlling subordinates. Only a decade ago hardly anybody had heard of benchmarking. Companies regarded it as an admission of defeat or a form of plagiarism. Now leading companies such as Xerox, Motorola, Hewlett-Packard and Ford routinely benchmark themselves against the best in the world to improve their performance. Internal benchmarking is also gaining ground. Analog Devices, a computer company, monitors the performance of all its units and regularly publishes reports on their progress. Internal benchmarking may be even more potent than external benchmarking because all the information needed will be readily accessible. Multinationals are better placed than smaller firms to carry out internal benchmarking because they have more variety to choose from.

Reprints
Reprints of this survey are available at US$3.75 each, plus 10% postage. Add tax in DC, IL, NJ, NY, VA; GST in Canada.

For classroom use or quantities over 25, please telephone for discount information.

Please send your order with payment by cheque or money order to The Economist Newspaper Group, Inc.,
Reprints Department,
111 W 57th St,
New York, NY 10019.
Telephone (212) 541-5730
(American Express accepted)

Large companies are also better placed to spread best practice through the use of internal competition. Many companies have star plants, such as Motorola's pager plant in Florida, which are held up as examples to everyone else. Mr Barnevik sends out a league table comparing the performance of different divisions to ensure that the worst performers learn from the best. AT&T encourages its telephone makers to compete with others in the group. One of its American plants, at Shreveport, recently managed to bring the cost of producing a cordless telephone down to the levels of the company's plant in Singapore.

Knowledge gained in one part of the world can be used to gain an advantage in another. Nynex, an American telecommunications company, is using state-of-the-art cabling techniques in virgin markets such as Thailand, in the hope of drawing on that experience when it comes to cabling its home market, the United States. Consultancies such as McKinsey make a handsome living by applying insider knowledge gained in one part of the world to a related problem somewhere else.

These companies are also becoming adept at using their multinational position to create knowledge from scratch. They can use internal competition to ensure that different bits of the organisation in different parts of the world vie with each other to come up with the best ideas. Digital, a computer company, encourages people in different parts of the organisation to compete to make new products. Toshiba, Fujitsu and a number of other companies hold international competitions to spur employees from all round the world to come up with the brightest idea or the most innovative new product.

If multinationals can turn themselves into global knowledge brokers, they may be entering a new, more optimistic era. For the past decade they have had a distinctly uncomfortable time as they tried to make amends for their past profligacy: dealing with bureaucratic bloat by sacking workers, delayering middle management and contracting out everything but their core functions to specialised companies. They have been bringing their operations up to date by encouraging their national affiliates to co-operate rather than duplicate, and trying to keep up with their smaller, nimbler rivals by breaking themselves up into federations of companies or rival profit centres.

Now, at last, these companies are starting to look ahead. The talk in the boardrooms is no longer just about downsizing and delayering, but about investing in "core competences"—the things that a company is best at doing—and shaping the market a decade hence. Companies such as ABB have shown that they can combine local focus with global reach. Companies such as Ford are trying to turn entire continents into production platforms for particular types of global cars. Companies such as Nynex are planning to redesign entire industries through a mixture of alliances, joint ventures and investments. The world's multinationals may not achieve all of their grand ambitions, but they have rediscovered the joys of thinking big. ■

RESEARCH
LAB
RULER
U-LIKE

67

WHARTON REPROGRAPHICS

15

Foreign Direct Investment as a Sequential Process

Bruce Kogut

The primary advantage of the multinational firm, as differentiated from a national corporation, lies in its flexibility to transfer resources across borders through a globally maximizing network. Recent models of direct foreign investment have tended to downplay these advantages of a co-ordinated multinational system; rather, they have stressed the motivational behaviour arising out of essentially national factors and market imperfections, e.g., proprietary knowledge, domestic industrial structure, and product differentiation. The neglect of the advantages of multinationality obscures, though, an important distinction between the original motivations to establish plants in foreign countries and the subsequent investment decisions. There is, in short, a fallacy of explanation of genesis in failing to distinguish between the initial investment decision and the subsequent incremental investment flows.

This paper argues that current foreign direct investment (FDI) must be understood as largely sequential flows stemming from the advantages of flexibility of a multinational system. The empirical foundation for this argument can be seen in the change over the past 30 years of the dominant channels of US FDI from new intercompany outflows to reinvested earnings. In 1970, the ratio of equity and intercompany account outflows to reinvested earnings was 1.39. By 1979, the ratio was 0.32. (The complete time series is given in Table 1.) The predominant share of FDI flows are incremental investments in already established subsidiaries. In light of these trends, previous theories of FDI which stress the oligopolistic behaviour of corporations in their home markets provide incomplete explanations for current FDI. What is required is a greater consideration of the systemic advantages inherent in a multinational network.

These trends carry implications also for the way data on FDI flows are categorized and used. Rather than collecting data according to entries and exits, information should be gathered concerning the conduit of flows as well as changes in the stock of FDI at the firm level. Data which reveals the industry and regional breakdown of both reinvested and

---

This reading has been published in *Multinational Corporations in the 1980s* (1983), Kindleberger, C. P. Audretsch, D. (eds.), MIT Press: Cambridge MA. Reprinted in *International Financial Management*, ed. Lessard, D. (2nd edn). Wiley: New York (1984).

B. Kogut, 'Foreign direct investment as a sequential process'. Reprinted with permission from *Multinational Corporations in the 1980s* by C. P. Kindleberger and D. Audretsch (eds)

Copyright © 1983 MIT Press, Cambridge, Mass.

Table 1. US Foreign Direct Investment (in millions of dollars)

Year	Total	Equity and intercompany account outflows	Reinvested earnings of incorporated affiliates	Ratio of equity outflows to reinvested earnings
1950	1,088	621	475	1.31
1955	1,766	823	962	0.86
1960	2,039	1,675	1,266	1.32
1965	4,994	3,468	1,542	2.25
1966	4,318	3,625	1,791	2.02
1967	4,768	3,050	1,757	1.74
1968	5,347	2,855	2,440	1.17
1969	6,186	3,130	2,830	1.11
1970	7,387	4,413	3,176	1.39
1971	7,280	4,441	3,176	1.40
1972	7,118	3,214	4,532	0.71
1973	11,435	3,195	8,158	0.39
1974	8,765	1,275	7,777	0.16
1975	13,971	6,196	8,048	0.77
1976	12,759	4,253	7,696	0.55
1977	13,039	5,612	7,286	0.77
1978	17,957	4,877	11,469	0.43
1979	24,844	5,904	18,414	0.32

Source: Survey of Current Business. US Department of Commerce (1981), 61(2).

new FDI flows would provide a critical platform by which the systemic advantages of the multinational corporation (MNC) can be appraised.

This is not to deny that the advantages of the MNC have been studied. The broader field of research on the MNC has been sensitive to these issues. Especially in the consideration of the political complexities posed by multinational corporations, the peculiar strengths attributed to the operation of a multinational network have been discussed.[1] Such issues as transfer pricing, tax arbitrage, bargaining or negotiating powers, and cost advantages have been analysed in isolation or in group.[2] There has not been, however, a thorough integration of these issues and the theory of foreign investment in the literature, with the partial exception of Dunning (1979).

The purpose of this paper is to move toward such an integration. Section 1 reviews briefly the recent FDI literature and introduces a more generalized model in the hope of elucidating the importance of the multinationality factor in FDI flows. The cutting edges of this model is the view of the MNC as a collection of valuable options which permits the discretionary choice of altering real economic activities or financial flows from one country to the next. Section 2 discusses the valuation of systemic advantages. Finally, Section 3 discusses the importance of this expanded theory from the point of view of host and home countries and comments upon trends likely to persist into the 1980s.

1. THE MULTINATIONALITY FACTOR IN FDI

Particular aspects of the contribution of multinationality to the value of the firm and to explaining its behaviour have been examined. Kindleberger (1969) considers the conflict between the host country and a multinational firm who are maximizing conflicting objective functions. Agmon and Lessard (1977) and Lessard (1979) note the incremental value of being able to arbitrage tax regimes. Hirsch (1976) cites in a revealing analysis the effects of joint production and trade in intermediate goods in the context of subsequent investment decisions. Vernon (1979) stresses the information- and profit-scanning functions of a multinational network. Davidson (1980) demonstrates the significance of experience effects upon FDI flows.

A growing body of research concerns the combination of location and trade theory with that of internalization. Magee (1977) argues that FDI is motivated by the difficulties of appropriating rents from the trade and licensing in proprietary knowledge. Buckley and Casson (1976) argue similarly that plant location is determined by, one, locational advantages and, two, market failure in the trade of proprietary knowledge, especially that of research and development.

Dunning (1977, 1979) has more recently expanded upon these ideas in developing what he calls the 'eclectic theory' of FDI. The theory combines usefully the macroeconomics of standard trade theory with transportation theory. Thus, a country's endowment and geographical position create certain 'locational' advantages. Dunning then proceeds to consider the factors which determine entry barriers and sustainable oligopolies. These factors, e.g. patented information, brand names, technology, form what he calls 'ownership' advantages. Ignoring the possibility of licensing, the various combinations of these advantages suggest a scheme such as shown in Figure 1. In the upper right box, for example, the firm possesses a unique technology or some cost advantage. Its home domicile is, however, characterized by higher factor or transportation costs than foreign locations. As a result, it invests overseas. Thus investment occurs only when the home firm possesses a unique asset and the host country is relatively advantaged in location. Finally, the theory of internalization is introduced to explain why licensing is not a preferred mode to FDI.[3] Thus, trade and location theory is wedded to that of internalization in order to explain FDI flows as a response to market imperfections.

		Location	
		Advantaged	Disadvantaged
Ownership	Advantaged	Export	FDI outward
	Disadvantaged	FDI inward	Import

Figure 1. Trade and FDI Outcome for a Single Industry in the Home Country.

Nevertheless, the importance of a multinational network as an important contribution to the value of the firm and its economic opportunities and to the determination of the likely conduits of FDI have not been sufficiently analysed. In the view of Buckley and Casson, for example, internationalization is the by-product of the search for minimum cost production sites and the internalization of markets. The absence of uniquely international factors is starkly apparent in Hirsch's model, which is considered in more detail in the next section. While Hirsch discusses economies of scale and joint production in marketing at a single location, there is no variable that represents the revenue impact of the ownership of a globally operated system. Dunning (1981) as well does not analyse the precise systemic opportunities generated by multinationality, though he does list several types of advantages.

But does the factor of multinationality add to our understanding of the determinants of FDI? One way to answer this question is to consider whether MNCs would still have cause to exist in the absence of the commonly listed market imperfections. Let us suppose, then, that the four of the five imperfections that Buckley and Casson list in the market for technology could be eliminated. These imperfections are (1) the absence of a forwards market to hedge the risk of development, (2) the impossibility to arrange contracts that would permit discriminatory pricing, (3) the presence of monopolistic purchasers, and (4) informational impactedness. (The fifth imperfection is that of government intervention, which—along with such events as changes in exchange rates—can create unique arbitrage opportunities for the MNC.) If in some idealized state, these imperfections were altered to the requirements of perfect markets, would there still be the MNC?

The answer would be affirmative, due to three characteristics which enable the MNC to exploit uniquely international distortions in markets or production: one, the ability to arbitrage institutional restrictions; two, the informational externalities captured by the firm in the conduct of international business; three, the cost saving gained by joint production in marketing and in manufacturing. These imperfections, which are carefully considered in the studies on the political dilemmas posed by multinationals and on multinational planning and control, are curiously understated in the mainstream economic literature.

This neglect is derived largely from not considering international projects as incremental at a global level. Although conceiving international investment within a dynamic theory of the firm, the internationalization literature fails to consider the value of the *operational* flexibility and externalities of a multinational system. Rather, the stress is placed upon the *structural* elements of plant location and the elimination of transactiona costs.

There is, in short, a tendency to view FDI as a decision made at a discrete point of time. This is the fallacy of genesis to which I previously alluded. The decision to transfer resources internationally is only one aspect of FDI. Given the structural outcome of this decision, the other aspect is the series of sequential decisions which determine the volume and direction of these transferred resources.

Contrary to the structural approach, consideration of the operational value of a globa system places stress upon the unique ability of multinationals to reduce the costs o operating in an uncertain world.[4] This is best illustrated by considering explicitly the three factors that were claimed above to be uniquely international attributes. The first is the ability of the MNC to arbitrage institutional restrictions, e.g., tax codes, antitrus provisions, financial limitations, and even national security prohibitions on trade. In effect, the operation of an international system has provided the multinational with a string of options written on contingent outcomes. The difference between the inclusion of taxes, financial incentives, etc., as part of a theory of location and this approach is that the consideration of institutional arbitrage as an option emphasizes the unique ability of the MNC to exploit the conditions of uncertainty and of institutional environments. The MNC can, in effect, exercise an option upon the occurrence of an event, e.g., its option to choose in which country to declare its profits. Boundaries do not represent only the costs of tariffs and transport; they also represent profit opportunities which can only be exploited by a multinational corporation.

The second factor concerns the capture of externalities in information, or what we call learning cost externalities. There is an information set required of international business that is separable from that required of domestic business. Corporations spend vast resources in their recruitment of internationally skilled personnel, in political analysis, in intercultural education programmes, and in the development of monitoring and control mechanisms. They also invest in information scanning and processing in order to locate markets and customers internationally. But most critically, there are important learning curves involved in these activities as well as 'first mover advantages'. Occidental Petroleum's knowledge and status as a long-standing supplier/customer of the Soviet Union is not easily duplicated.

Joint production economies can occur in both marketing and manufacturing on a global scale. Niehans (1977) has noted that economies of scale in marketing or servicing but constant or decreasing scale economies in manufacturing can serve to explain the growth of the MNC. Similarly, joint production economies due to the creation of a multinational network reduces the physical capital or labour costs of production and marketing of incremental investments. For example, the multinational network permits the export of otherwise non-exportable goods, since the fixed costs of establishing sales offices, hiring personnel, and locating plant sites are already sunk. The incremental cost saving can permit, as Hirsch suggests, an increase in the export of intermediate products or in the market entry of new products sharing production economies. In addition, the multinational network can serve to export additional final goods or to service the export of other firms' goods in times of slack capacity; the Japanese trading companies are an example of the latter facility. Unquestionably, the formation of a multinational network poses significant barriers to entry. The implications of these barriers are analysed in greater detail in the conclusions.

2. SYSTEMIC ADVANTAGES

The implications of considering the value of international factors can be illustrated in the formulation of a capital-budgeting procedure. In an instructive article, Hirsch (1976) suggests that the decision to service a market by exports or by foreign direct investments can be modelled by four cost variables discounted at some appropriate rate: production and tax transportation costs P, research and development costs K, marketing costs M, and control costs C. Demand is assumed as given. Marketing costs are argued to be higher for exports than host-manufactured goods, while control costs rise with the internationalization of production. These assumptions are intuitively reasonable. Moreover, Hirsch argues that M and C are increasing functions of K. Implicitly, Magee argued the same in his claim that the opportunity loss of appropriability rises with information content at every stage of production. Hirsch's reasons rest in the costs of market exploration and organizational communication. Since neither argument is mutually exclusive, both can be subsumed in Hirsch's model.

Given this formulation of the model, Hirsch demonstrates convincingly that the failure of conventional trade theory is its consideration solely of the comparative costs of production. Expanding the model to include the production of several goods, Hirsch also shows that economies of joint production can change the investment decision when production for each good is maintained as an independent project. The model is also expanded to consider multistage production, whereby Hirsch shows that overseas investment production can increase the export of previously non-competitive intermediary goods.

While Hirsch discusses economies of joint production in marketing at a single location, there is no variable that represents the cost savings of the ownership of a globally operated system. As suggested earlier, these cost savings arise from considering a global network as a sunk cost and sequential investments as incremental. (There is also the factor that in a world of imperfectly diffused information, the scanning advantages of a multinational widens its set of investment opportunities.) The impact of these cost factors is to reduce the incremental values of *C*, *K*, and *M* and to expand the firm's investment opportunities.

We can illustrate the value of systemic advantages by considering explicitly how a capital budget formulation might account for these benefits, particularly those of the first factor. The first factor, that of arbitraging institutional and national barriers, contributes directly to the enhancement of the discounted revenue stream. Such a contribution is derived through, one, a possible reduction in the discount rate, and secondly, through the addition of a string of real but usually non-tradeable options. The advantages accrued by international diversification in financial markets and the role of MNCs in providing diversification have been commented upon by several authors, in particular by Sonlik (1974), Lessard (1976), and Agmon and Lessard (1977). In a world of no barriers to portfolio flows and of purchasing power parity, the cost of capital at the margin is the same for all investors in a taxless world, regardless of nationality. Since changes in the exchange rates and nominal interest rates would reflect equally changes in prices, discounted cash flows before taxes are the same no matter what their currency denomination.[5] When capital markets are segmented by national barriers and policies, discount rates vary between countries. These variations reflect national differences in risk bearing and in time preferences. To the extent that barriers are imperfect, discount rates tend toward a world rate plus or minus the transactional costs of subventing these barriers. Considered in this light, the equilibrium discount rates can be viewed as prices for a single commodity adjusted for transportation costs, when all countries are equally distant from one another.

The argument for the advantages of the MNC in financial markets boils down to the following set of simple statements. Multinational firms are able to invest across national borders and thereby avoid the presumably more costly barriers to portfolio flows. Because of its access to more diversified international financial markets, the MNC can, holding everything else equivalent, invest in marginal projects otherwise rejected by host firms. Since the MNC adjusts its risk according to a world portfolio and the domestic firm according to a more home-weighted portfolio (due to capital market imperfections and barriers), the risk-adjusted cost of capital is relatively lower for the MNC. Secondly, if the foreign project represents a unique asset in terms of the available portfolio investments on the world market, the home investor is willing to pay the MNC a premium for contributing an otherwise unattainable asset to his or her portfolio. In conclusion, then, in a world of imperfect international capital flows, MNCs receive a premium relative to their purely domestic home competitors—if their foreign investments are unique and non-traded assets—and invest at a lower capital cost relative to the purely domestic host country competitors.[6]

An aspect that has received less attention in the literature regarding the evaluation of a multinational network is the value of holding a string of options defined by institutional barriers. These barriers can be typified by different currencies or more exactly by changes in relative exchange rates, taxes, sovereign risk, and legal prescriptions on equity ownership, form of remittance, etc. To illustrate the importance of these options, consider an MNC operating a series of plants in several countries which serve purely domestic markets. The future cash flows are uncertain. Upon realization of its cash flows, the MNC can through transfer pricing, financial packaging, and other methods alter to a non-trivial extent the structure and level of its obligations. The contingent events are the realization of taxable earnings in each of the various countries. The nature of the option is that the MNC can choose in which geographical jurisdiction to declare these profits so as to minimize the tax burden.

Consider, for a second example, the MNC with several export markets and a few manufacturing plants. By design, the plants operate on average at less than full capacity. In country *X*, it must renegotiate a set of extraction or labour contracts. In response to unacceptably (however subjectively defined) bargaining positions, the MNC can shift production to other plants to service its overseas markets, given that labour constraints are not binding. (This potential has secondary effects in terms of its bargaining strength, as discussed in Section 3.) Thus given the uncertainty over the outcomes of negotiations, the MNC has the real option of mitigating the costs of factor price increases or higher taxes in the short term, and the physical adjustment costs in the long term, through maintaining excess capacity in diverse national plants. A similar case is that of a real appreciating currency. The ability to shift exports from a country whose currency is appreciating to one where other plants are located is a valuable option—assuming that pricing is denominated in the currency of the importing country and factor payments are derived out of overseas revenues.

What is important to note is that all the above-described options are valuable because the future state of the world is uncertain.[7] Moreover, the more variable the environment, the more valuable are these options. From the point of view of the MNC, the variance of its cash flows and of factor prices in the context of national restrictions presents a set of valuable options relative to the opportunity set of a purely domestic firm. Since these options are exercisable only by the MNC and cannot be traded and purchased by individual investors in any meaningful sense, the value of the firm is enhanced by the incremental value of these options.[8]

The above discussion can be illustrated through the consideration of the elements that would enter into a capital-budgeting model. The value of the firm can be described as

$$NPV=\sum_{0}^{t}\frac{(\text{cash flows} + \text{learning} + \text{joint production} + \text{options})}{(1+R)}$$

The discount rate may be lower than that available on the world market and, as is more likely, lower than that available to host country firms when capital markets are segmented. (The appropriate discount rate for each term is likely to vary according to its systematic risk, but for simplicity, we assume one rate.) The first term represents simply the discounted cash flows from a series of independent projects that are owned by globally unintegrated firms. The next three terms capture the advantages arising out of the interdependence of the cash flows of projects undertaken by an MNC. We have indicated these advantages to be learning-cost and joint-production economies, and the possession of real non-tradeable options.

How may these advantages be actually evaluated?[9] Once the initial investment is made, learning cost and joint production externalities are likely to be incorporated in the capital budget through an incremental analysis. A more complex analysis is to evaluate at the time of the initial investment the value of the option to expand into other products or into other markets. The option to arbitrage national barriers is similarly difficult to evaluate. Given the recent interest in evaluating financial options, it would appear that the analytical methods are already available to value real options numerically.[10] There are, however, serious obstacles to such an extension which are not always noted in the literature. In brief, a financial option can be valued because one assumes that there exists a shadow security whose price will follow a particular type of stochastic process. In the absence of such a shadow security, we have no reason to believe that the price of the real asset will also follow such a path since there is no means to arbitrage the real option against its shadow security. Nevertheless, the options literature provides the appropriate framework for understanding the reasoning behind the claim that the option to exercise certain rights, e.g., where to declare taxes, where to shift production, is a valuable hedge against contingent events.

The inclusion of these three elements in addition to the normal cash flows generated by project illustrates the opportunities that stem from a multinational system. As a result of these opportunities, we expect that growth in FDI is more likely to be in the form of reinvested earnings than in new entries. The consequences of this trend from the perspective of governments are discussed in the next section.

3. CURRENT TRENDS

From the firms' point of view, the flexibility to transfer resources across borders is a positive contribution to its earning stream. As discussed in an abundant literature, the viewpoints of governments are often less sanguine. Rather than consider the larger issues of government and MNC relations, we concentrate on a few specific issues, e.g., monetary stabilization, regulation, and negotiation. Through these examples, we illustrate that the well-established issues of contention between governments and the MNCs are comprehensible only within the context of the systemic flexibility of the MNCs.

How is a government likely to view these systemic advantages of MNCs when pursuing domestic monetary stabilization objectives? If an MNC speculates on currencies, its behaviour tends to give markets greater liquidity and thereby accelerates the speed of adjustment. Even if the MNC is not a speculator, it forecasts and hedges its contractual and non-contractual exposure. Whether it speculates or is able merely to shift currencies more easily from one country to the next or to write more inexpensive contractual hedges, these activities clearly tend to negate the ability of governments to pursue objectives such as the stabilization of its currency countervailing to the market trend. Thus the presence of MNCs limits further the ability of governments to pursue independent objectives in an international economy.

If governments are more constrained in their ability to pursue monetary objectives due to the existence of MNCs, are they more constrained in pursuing other objectives? Consider a government which desires to regulate a foreign-owned industry. Under closed borders, enterprises are essentially hostage to the partisan coalitions of its environment. Since exit except in the form of selling its holdings is impossible, the regulated enterprise consents to government regulation, while it may itself seek to join a political coalition.[11] For the MNC whose foreign assets are primarily in the form of proprietary knowledge, regulation is unlikely to be successful. It is especially unsuccessful in its most extreme form—that of nationalization, for as long as the ongoing value of the subsidiary is dependent on a sequential stream of innovations, nationalization results merely in the elimination of industry. The effects of less extreme interventions, e.g., an increase in taxation, invoke most likely less extreme reactions. Nevertheless, the implications are the same. To the extent that multinational corporations contribute more to the economy than that of the new policies, their partial or total withdrawal are a real loss to the host country. Moreover, if multinational corporations are truly global, then the argument applies to home countries in which governments attempt to regulate, for example, outward flows of capital.

When FDI is primarily transmitted in the form of fixed capital with known and stable technologies, the bargaining position of the MNC and host government is reversed. This situation, which is known as the obsolescent bargain, has been extensively analysed.[12] Mining is a classic example. The original investment requires a large fixed capital component. Presumably the MNC chooses to internalize this transaction in order to gain access to guaranteed supplies, to speculate on future prices when futures markets do not exist or do not trade claims on contracts many years out, or to acquire oligopolistic advantages. At the time of the investment, the host government's bargaining position is constrained by competition from the other countries. The host government can, however, capture part, if not all, of the rents ex post to the investment through a process of renegotiation. If, for example, the value of its mineral resources is higher than expected, the host government has the option of capturing part of the consequent excess returns. To the extent that this gaming behaviour is expected and that equity claims and other financial contracts are difficult to enforce, it is likely that MNCs will be reluctant to invest in mineral extraction without adequate guarantees or insurance.

What does the above analysis imply for the growth and the role of the MNC in the 1980s? There are three principal implications. First is that the entry of new firms in international markets is likely to slow, holding changes in investment opportunities constant. Evidence for this transition is suggested, as noted earlier, in the growth of reinvested earnings relative to new equity investment in total FDI flows. These figures have been given in Table 1.

The benefits of institutional arbitrage, learning curve effects, and joint production economies create, then, substantial barriers to entry. The importance of these factors is underlined in light of the present proportion of intra-firm trade which accounted roughly for 48.4% of all US imports in 1977 (Helleiner, 1979). Discussions of antitrust implications have usually stressed the feedback of FDI on competition in the home market (Bergsten et al, 1978). Perhaps a more troubling aspect is the impact of these barriers on the entry of firms from LDCs or from developed countries which were slow to create international firms. While FDI between LDCs has been increasing, such investments appear to be characterized more by the capture by small firms of investment gaps ignored by the much larger MNCs. Thus, FDI by LDCs is likely to be explained by a theory of FDI by small firms than by entry openings unique to the skills of LDC enterprises.[13]

Another implication lies in the persisting role of research and development in explaining some, though perhaps a relatively decreasing amount, of flows of proprietary capital over international borders. For those end products that are oriented towards the home market, FDI is often in the form of overseas production of intermediary products. In this case, FDI is related to research and development only insofar as the end product embodies a large value of technological expenditure. The overseas production of intermediary goods can involve minimal research and development expenditures in terms of value added. Moreover, FDI is in this form partially trade enhancing.

What should be noted is that the production of technologically sophisticated products is of less importance—though still of indubitable significance—in future FDI flows. Instead, we can imagine the development of decentralized structures that permits the delegation of product selection and research and development to subsidiaries but leaves the strategic variables of financing, production, and tax arbitrage to the home office. There is some indication that this devolution is already perceptible in the food industry (Katz, 1981).

Related to the above implications is the tendency of the MNCs to develop and expand their trading divisions. In part, this evolution is derived from the hazards of the obsolescent bargain as well as from the erosion of market advantages as the original technological edge in some products evaporates. Vernon (1977) has termed this latter trend 'senescence'. The difference between these two trends is simply that the obsolescent bargain refers to loss of property rights, senescence to loss of market shares. The underlying cause of both trends is, however, similar, i.e., the loss of some technological advantage specific to the firm which maintains its bargaining or market position. In the first case, the MNC can limit its exposure by financing extraction in the form of debt with payoffs denominated in a specified quantity of the underlying mineral or product, or by long-term contracts, or by providing managerial services and downstream marketing and relinquishing its ownership or contractual obligations. In the second case, the MNC is also induced toward eliminating its productive activities and in effect leasing the services of its global network. In both cases, the possession of a multinational network provides a stream of benefits and investment opportunities independent of the products being traded. Examples of this development are the oil industry, Japanese trading companies, and the diversification of large firms into trading third-party products, such as in the case of Thyssen.

Consequently, the MNC of the 1980s is likely to be engaged less in equity investments in primary extraction industries, but relatively more in the provision of marketing and consulting services. Trade in intermediate products is also likely to increase, because the MNC can optimize production and marketing within an already existing global system. These trends, of course, are a continuation of a pattern that has been visible for several years.

4. CONCLUSIONS

I have tried to detail the precise advantages arising from a multinational network and its implication for the identity of the agents and the type of investment flows in the future. By and large, policy implications have not been discussed. The reasons for this omission are simple. Having developed our model under the assumption of profit maximization, optimal policy recommendations which differ from a competitive profit-maximizing outcome can only be motivated if the national objective functions are specified. I have not attempted such a specification, although the dropping of the assumption of profit maximization would be an interesting exercise.

The question has been left open whether the profitability and growth of the MNC is a result of its market power or productive efficiency.[14] I have, however, implicitly suggested the importance of first-mover advantages that current MNCs possess relative to potential entrants. If we are concerned over the absence of bargaining power on behalf of many LDCs, then our analysis reinforces the recommendations for an international regulation of MNCs or the creation of countervailing institutions or enterprises to enhance the bargaining power of LDC countries. Such regional efforts as ASEAN's recent consideration of the formation of trading companies similar to those of the Japanese are illustrations of efforts to create potential countervailing enterprises. But hidden beneath such developments and recommendations is an irony often noted in the case of MNCs from developed countries. That is, though the rents from learning-curve and joint-production externalities accrue to the home MNC and hence potentially to the home countries, the creation of truly global enterprises poses challenges to the national sovereignty of governments through their maximization of global profits and through their arbitraging of institutional borders. In other words, is it reasonable to expect that an MNC originating in a developing country, which MNC maximizes its return from its global activities, should be more sensitive to the sovereignty and interests of its national government?

Consequently, the conflict of nation-states and international firms remains an issue in the 1980s. Countries face incentives not only to regulate the entry of firms, but also their exit. There has appeared the ironic evolution that LDCs have been concerned to establish strict rules of entry, whereas developed countries increasingly seek to control the exit of firms and the immediate loss of jobs and production. Whether the combined impact of these trends is to reduce the benefits of a global network remains to be seen; but, in any event, it represents the forefront of future discussions on the merits of FDI and the multinational corporation.

NOTES

I would like to thank Stephen Kobrin of New York University and Richard D. Robinson of the Massachusetts Institute of Technology for their comments on an earlier draft. I am especially grateful to Donald Lessard of the Massachusetts Institute of Technology for his comments on the first and subsequent drafts.

1. See, for example, Kindleberger (1969), Vernon (1971, 1977), or Stopford and Wells (1972).

2. Such issues as transfer pricing, tax arbitrage, bargaining or negotiating power, and regulation are discussed by Lessard (1979), Bergsten et al (1978), and Robinson (1976).

3. For a review of the internalization literature, see Rugman (1981). It should be noted that the concept of internalization is already adumbrated, like so many ideas in the literature on FDI, in Kindleberger (1969), pp. 19–22.

4. Undoubtedly, it can be claimed that the theory of internalization accounts precisely for these operational facets. (See, e.g. Buckley and Casson, 1976, p. 69.) But by placing stress upon the cost aspect of transactions, it fails to consider the profit opportunities generated by a global system.

5. This can be shown by considering whether to discount the overseas earnings by the home or foreign discount rate. If we assume purchasing power parity, changes of nominal interest rates are cancelled by identical changes of nominal exchange rates.

6. Robert Aliber (1983) has argued that FDI can be explained at the macroeconomic level, i.e., the arbitrage of international markets. Though such an approach appears weak in explaining such phenomena as cross-hauling, it has a tantalizing appeal on its attempts to link such anomalies as FDI waves to concepts as Tobin's *q*. There may be a macroeconomic story after all.

7. In the cases where uncertainty may appear as irrelevant, such as in tax arbitrage, governments are well equipped to develop monitoring and enforcement services. It is the uncertainty of the realized profits that keeps the costs of these services relatively high to the benefits of reducing arbitrage behaviour.

8. Some readers may be misled into inferring that the above argument suggests that total variance does matter after all to the investor. To the contrary, since these options are similar to monopoly or proprietary rents, the firms earn abnormally high rates of return but the stocks written on the firm, as long as they are traded in competitive capital markets, are priced in the expectation of a market rate of return adjusted for systematic risk.

9. See Lessard (1981) for a thorough discussion of the application of adjusted present value techniques for evaluating international projects.

10. See the path-breaking article by Black and Scholes (1973) for the evaluation of financial options. Our concern with the shadow security is directed primarily at recent extensions of the Black–Scholes model into the evaluation of real assets. MacDonald and Segal (1981) are certainly aware of these difficulties, whereas Cooper and Broglie (1981) simply assume the existence of a shadow security without substantial comment.

11. See the interesting article by Magee in this volume which discusses the impact of political coalitions on FDI flows.

12. See the writings of Vernon (1971), Stopford and Wells (1972), and Bergsten et al, (1978).

13. We have not tried here to develop such a theory of FDI by small firms. Briefly, one relevant factor would seem to be trade in custom-designed products, i.e., small producers are more sensitive to the 'voice' (in Hirschman's terminology) of smaller producers (Hirschman, 1971). Another factor is the trading by small firms of used capital equipment. Since used capital equipment is difficult to evaluate, the seller may attempt to eliminate the costs of discounting incurred through the asymmetry in information by taking an equity position. Indeed, joint ventures between LDCs are relatively common. These factors, in addition to the ones discussed above, tend to explain some characteristics of FDI between LDCs, whose markets tend after all not to be dominated by large domestic enterprises as those in developed countries. For a discussion of FDI between LDCs, see Wells (1977).

14. A number of articles in *Multinational Corporations in the 1980s* attempt to measure econometrically the relationship between the profitability of the MNC and barriers to entry. Though the factor of efficiency is not explicitly specified in these regressions, the results are nevertheless interesting in showing the significant correlation of these profits and barriers to entry.

REFERENCES

Aliber, R. (1983) in *Multinational Corporations in the 1980s*, Kindleberger, C. P., Andretsch, D. (eds). MIT Press: Cambridge, MA.

Agmon, T., and Donald L., 1977. 'Financial Factors and the International Expansion of Small-Country Firms,' in Agmon, T. and Kindleberger, C. P. (eds.), *Multinationals from Small Countries*. Cambridge, Mass.: MIT Press.

Bergsten, C. F., Horst, T., and Moran, T. (1978). *American Multinationals and American Interests*. Washington, DC.: Brookings.

Broyles, J. E. and Cooper, I. A. (1981). 'Growth Opportunities and Real Investment Decisions,' in Derkinderen, F. G. L. and Crum, R. L. (eds.), *Risk, Capital Costs, and Project Financing Decisions*. The Hague: Nijhoff.

Buckley, P. J. and Casson, M. (1976). *The Future of the Multinational Enterprise*. London: Holmes and Meier.

Davidson, W. (1980). 'The Location of Foreign Direct Investment Activity: Country Characteristics and Experience Effects.' *Journal of International Business Studies*, pp. 9–22.

Dunning, J. H., (1977). 'Trade, Location of Economic Activity and the MNE: A Search for an Eclectic Approach,' in Bertil Ohlin et al, (eds.) *The International Allocation of Economic Activity*. London: Holmes and Meier.

——, (1979). 'Explaining Changing Patterns of International Production: In Defense of the Eclectic Theory,' *Oxford Bulletin of Economics and Statistics* 41, pp. 269–96.

——, (1981). 'Explaining Outward Direct Investments of Developing Countries: In support of the Eclectic Theory of International Production,' in *Multinationals from Developing Countries*. Lexington, Mass.: Maxwell G. McLeod.

Helleiner, G. K. (1979). 'Transnational Corporations and Trade Structure: The Role of Intra-firm Trade,' in H. Giersch (ed.), *On the Economics of Intra-Industry Trade*. Tübingen: J. C. B. Mohr.

Hirsch, S., (1976). 'An International Trade and Investment Theory of the Firm,' *Oxford Economic Papers* 28, pp. 258–270.

Hirschman, A. O. (1970). *Exit, Voice, and Loyalty, Responses to Decline in Firms, Organizations, and States*. Cambridge, Mass.: Harvard University Press.

Katz, J. H. (1981). 'Does Foreign Direct Investment Theory Reflect Reality: The Case of the American Food Processors.' Unpublished paper, October: to be published as a Sloan School of Management Working Paper. Massachusetts Institute of Technology.

Kindleberger, C. P. (1969). *American Business Abroad: Six Lectures on Direct Investment*. New Haven: Yale University Press.

Lessard, D. (1976). 'World, Country, and Industry Relations in Equity Returns: Implications for Risk Reduction through International Diversification,' *Financial Analysts' Journal*. January, February, pp. 2–8.

—— (1979). 'Transfer Prices, Taxes and Financial Markets: Implications of Internal Financial Transfers Within the Multinational Firm,' in R. G. Hawkins (ed.), *Economic Issues of Multinational Firms*. New York: JAI Press.

—— (1980). 'Evaluating International Projects: An Adjusted Present Value Approach.' in Crum, R. and Derkinderen, F. (eds.), *Capital Budgeting Under Conditions of Uncertainty*. The Hague: Nijhoff.

McDonald, R. and Segal, D. (1981). 'Options and the Valuation of Risky Projects' (unpublished paper), August, presented at the Finance Research Seminar of Sloan School of Management, Massachusetts Institute of Technology.

Magee, S. P. (1977). 'Information and the Multinational Corporation: An Appropriability Theory of Direct Foreign Investment.' in J. Bhagwati (ed.) *The New Economic Order: The North–South Debate*. Cambridge, Mass.: MIT Press.

Niehans, J. (1977). 'Benefits of Multinational Firms for a Small Parent Economy: The Case of Switzerland,' in Agmon, T. and Kindleberger, C. P. (eds.) *Multinationals from Small Countries*. Cambridge, Mass.: MIT Press.

Robinson, R. D. (1976). *Natural Control of Foreign Business Entry*. New York: Praeger.

Rugman, A. M. (1980). 'Internalization as a General Theory of Foreign Direct Investment: A Re-Appraisal of the Literature.' *Weltwirtschaftliches Archiv* 116, pp. 365–79.

Solnik, B. H. (1974). 'Why Not Diversify Internationally? *Financial Analysts' Journal*. July–August, pp. 48–54.

Stopford, J. and Wells, L. (1972). *Managing the Multinational Enterprise Organization of the Firm and Ownership of the Subsidiaries*. New York: Basic Books.

Vernon, R. (1971). *Sovereignty at Bay: The Multinational Spread of US Enterprises*. New York: Basic Books.

——, (1977). *Storm over the Multinationals: The Real Issues*. Cambridge, Mass.: Harvard University Press.

—— (1979). 'The Product Cycle Hypothesis in a New International Environment,' *Oxford Bulletin of Economics and Statistics*, 41, pp. 255–267.

Wells, L. (1977). 'Firms from Developing Countries,' in Agmon, T. and Kindleberger, C. P. (eds.), *Multinationals from Small Countries*. Cambridge, Mass.: MIT Press.

911
673
68
WHARTON REPROGRAPHICS

LIPPINCOTT LIBRARY
DOCUMENT DELIVERY SERVICE

Global or Stateless Corporations Are National Firms with International Operations

Yao-Su Hu

The debate on American competitiveness in the world economy has recently been thrown into some disarray. Robert Reich asserts in his essay "Who Is Us?" that in a "borderless" world characterized by "stateless" corporations, it makes no sense to talk of American competitiveness in terms of American-owned corporations, but rather that American competitiveness should be defined in terms of the skills and experience of the American work force.[1] And as Kenichi Ohmae states: "It does not matter who builds the factory or who owns the office building or whose money lies behind the shopping mall or whose equity makes the local operation possible. What matters is that the global corporations . . . act as responsible corporate citizens."[2]

Since these aggressive assertions hinge on the concept of the "global" (or "stateless") corporation, the term deserves careful scrutiny. It has been preceded in the last 30 years by a string of synonyms such as: international, inter-territorial, multinational, transnational, and worldwide. The concept has evolved to the point where it connotes something that has gone beyond nations and has left them behind. Before this, economists talked of "direct foreign investment," which implied the existence of a home nation and "host" (i.e., foreign) countries. In 1960, David E. Lilienthal, who had achieved prominence as Director of the Tennessee Valley Authority and then as Director of the Atomic Energy Commission, launched the expression "multinational corporation."[3] Since then, exaggerated claims have been made about the nature of this newly discovered species and about the inevitable demise of the nation-state. MIT economics professor Charles

I would like to thank Wyn Grant, John Hutton, Dan Jones, Keith Pavitt, Sol Picciotto, Anthony Steele, and Richard Whitley for comments and suggestions.

Reproduced with permission of copyright owner. Further reproduction prohibited.

Kindleberger wrote in 1969: "The international corporation has no country to which it owes more loyalty than any other, nor any country where it feels completely at home . . . The nation-state is just about through as an economic unit . . . The world is too small. It is too easy to get about."[4] What differs between the 1960s and the 1990s is the context of the assertions: in the 1960s, it was the Europeans who worried about U.S. multinationals; now it is the Americans who worry about Japanese and European multinationals.

This article addresses some difficult and fundamental questions.[5] What is the nature of the "global," "transnational," or "multinational" corporation? Is it "national" (albeit with foreign or international operations) or is it "stateless" or "global" (by which it is meant that the company has really transcended nations in the sense that it is indifferent as between different countries and has no home nation)? What are the strategic, overall implications for the firm's competitive position in the world economy? And finally, what are the analytical, conceptual implications—i.e., what difference does it make to the way in which we think about the issues, to public debate, and to the research agenda?

Nationality Versus Stateless Globalism in the MNC

The argument that MNCs have become stateless entities stems from the observation that their business operations cross national borders. General Motors makes the Pontiac LeMans in Korea to sell in the U.S., Honda makes cars in Ohio that it exports to Japan, and Boeing makes sections of its planes in China, Italy, and Canada for assembly in Seattle.[6] However, this argument is questionable on two counts. First, stateless operations do not necessarily mean stateless corporations; in addition to the geographical spread of the group's operations, there are other criteria that need to be considered before one can consider the group to be stateless—criteria such as the ownership, control, top management, and legal nationality of the group and its components. Second, it does not necessarily follow from the fact that operations cross national boundaries that the nations are of equal importance to the group or that there is no geographical center of gravity.

Some schools of thought label a firm as a "global" or "transnational" corporation simply because it adopts a "global" or "transnational" strategy. Earlier, it was the adoption of a global habit of mind that distinguished the "geocentric" company from the "ethnocentric" and the "polycentric" firm.[7] Apart from the fact that a global strategy has meant different things to different writers, the problem is that a strategy or a habit of mind cannot be quantified. Nor can it be measured unambiguously through its results, for the actual outcome depends not only on the strategy but also on the interplay of internal and external forces.

Geographical Spread and Scope—For a company to be stateless in the sense of being indifferent between nations, a necessary condition is that its

Reproduced with permission of copyright owner. Further reproduction prohibited.

operations should be evenly distributed among these nations. This raises the question of what indicators should be used to measure the firm's operations. It seems clear that the distribution of sales or turnover cannot by itself be a satisfactory indicator—for example, a company may be deriving 80% of its total sales from foreign countries through exporting, without any foreign manufacturing, yet it would remain clearly anchored in its home nation. The distribution of profits is even more problematic, as this not only depends on the distribution of sales, but is further affected by inter-country transfer pricing.

The geographical distribution of value-added (or net output—i.e., the value of output minus the value of purchased inputs) would be an ideal indicator, but I have not come across a single multinational whose published accounts provide enough data for this to be computed. We therefore have to fall back on the distribution of a firm's assets (total or fixed assets) and of its number of employees as approximate indicators. Table 1 illustrates this for a few of the best-known U.S. multinationals. Many of the German and Japanese multinationals also have less than 50% of their operations abroad. On the other hand, many of the British MNCs and MNCs from the smaller European nations such as Switzerland have more than 50% of their operations located abroad; this is because of historical reasons or because of economic necessity (i.e., the need to reap economies of scale and scope). At the extreme is Switzerland's Nestlé, which has 95% of its assets and 96.5% of its employees located outside the home nation.

In cases where less than half of a group's operations are abroad, the home country operations are of overwhelming importance to the MNC. Such a firm cannot escape the influence of the home nation's environment. If it is based in a stagnant, declining, or otherwise unfavorable nation, it will be at a disadvantage in relation to competitors based in more dynamic nations. For example, imagine the case of two electronics companies, one American and one Japanese, both of which have 60% of their total operations in the home nation. If the Japanese electronics industry grows at 10% a year compared to 5% a year for the U.S. industry, and if each firm

Table 1. The International Scope of Some Multinationals

Company	Percent of total assets outside home country	Percent of total no. of employees outside home country
IBM (1989)	46	44
General Motors (1989)	24	31
Du Pont (1989)	35	24
General Electric (1989)	9	17
Sample Nonbank U.S. MNCs (1988)	22	26

Sources: Computed from company annual reports; U.S. Department of Commerce, Survey of Current Business, June 1990.

Reproduced with permission of copyright owner. Further reproduction prohibited.

maintains its share of its national industry, the Japanese firm will have a higher overall rate of growth than the U.S. firm if the overseas operations of both grow at the same rate. If the overseas operations grow at 10% per annum, the Japanese firm will grow at 10% overall, the U.S. company at 7%. If the U.S. firm wants to match the Japanese rate of growth, it has to achieve a higher rate of growth of its overseas operations than the Japanese firm achieves on its overseas operations; thus if the Japanese firm's overseas operations grow at 10%, the U.S. firm's foreign operations have to grow at 17.5% if it is to attain the same overall rate of growth (10%). This is a tall order, especially if the U.S. firm has to defend its home market against Japanese competition at the same time. This dependence of the corporation's fortunes on the home nation, depending on the relative weight of home operations, results from the fact that the operations of the global enterprise have a center of gravity or home nation.

What if the group's foreign operations exceed 50% of the total? It may be thought that this would put foreign operations on the same footing as domestic operations in the eyes of the headquarters. However, although overseas operations exceed 50% in aggregate, this 50% is divided between a number of foreign (host) nations. Thus, any individual host nation is likely to account for a much smaller percentage of the corporate total than the home nation. This is illustrated in Table 2.

ICI's case is typical of most MNC's in the world, regardless of their national origins: the home nation is more important, in sheer quantitative terms, to the group than any single foreign country. Furthermore, since governments, civic and business associations, labor unions, and pressure groups are organized and act most effectively within the limits of the nation, the global firm is more susceptible to pressure, persuasion, or requests for cooperation coming from the home nation than from any other country. To put it simply, the home government has within its jurisdiction a much bigger chunk of the group's assets than any individual foreign government. Given that the MNC is a political as well as an economic phenome-

Table 2. International Spread of Imperial Chemical Industries, 1989

	Percent of total net operating asset	Percent of total no. of employees
Overseas operations	62.6	59.1
UK (home base)	37.4	40.9
Continental Europe	15.4	12.5
The Americas	27.5	25.3
Asia Pacific	15.8	12.5
Other countries	3.9	8.9

Source: company annual report

Reproduced with permission of copyright owner. Further reproduction prohibited.

non, this differential susceptibility and receptivity is of great importance to the nations concerned. It often is not true, as Kindleberger stated, that the MNC has no country to which it owes more loyalty than any other: self-interest dictates otherwise. It should be noted that the issue here concerns the MNC's differential susceptibility; the fact that the home government may choose not to exert pressure is another matter.

There are two exceptions this rule. First, there are companies from small nations, companies like Nestlé, for which a single foreign country (e.g., Germany) may represent a higher percentage of total operations than the home nation. Second, there are "binational" companies which have two home nations or centers of gravity: Shell and Unilever are both based in the UK and the Netherlands, and Asea Brown Boveri is based in Sweden and Switzerland. These two exceptions represent, in this day and age, the most advanced form of internationalization of the corporate entity.

Ownership and Control—A global enterprise is normally made up of a parent company, located in the home nation, and a number of subsidiaries in host nations. A subsidiary is like any local company in that it is a legally incorporated entity and has its own legal identity; its distinctive characteristic is that it is owned and/or controlled by another corporation. A subsidiary can be wholly or partly owned by the parent company. Nowadays, partly owned subsidiaries usually involve joint ventures with another parent company or situations in which host-country laws require the presence of local shareholders. There seems to have been a tendency on the part of many MNCs to eliminate shareholdings by the public in their overseas subsidiaries, holdings which owed their existence to historical legacy. In 1959, IBM bought out the 38% minority holdings in its UK subsidiary and in 1961 Ford bought out the 46% minority shareholdings in Ford England.[8]

At the level of the parent company, ownership and control remain national rather than multinational. Although foreigners may owns shares in the publicly quoted parent company, in most cases the majority of the shares are held by individuals and legal entities from the home nation.[9] Control naturally follows, so that both top management and governance rest in national hands. Thus, in terms of ownership and control, there is no doubt that a company like Siemens is German, a company like IBM is American, or a company like Toyota is Japanese. In the case of Nestlé (which has less than 5% of its total assets and employees in Switzerland), it is noteworthy that Swiss law allows Swiss companies to exclude foreigners from holding registered shares (which alone carry voting rights) and that Nestlé limits non-Swiss voting rights to 3% of the total.[10] In this respect, even a firm such as Nestlé is firmly national rather than transnational or stateless.

The furthest extent to which a group's ownership and control have been internationalized is the case of the "binational" companies, in which the

Reproduced with permission of copyright owner. Further reproduction prohibited.

group has two parent companies. The Royal Dutch/Shell Group is 40% owned by the Shell Transport and Trading Company PLC (UK) and 60% owned by the Royal Dutch Petroleum Company (Netherlands); the two parent companies are owned mainly in their respective home nations. Similarly, the Unilever Group has two parents, Unilever N.V. and Unilever PLC, and Asea Brown Boveri is owned, in equal parts, by ASEA AB (Sweden) and BBC Brown Boveri Ltd (Switzerland).

The multinationalization of shareholdings in the parent companies of global enterprise groups cannot be ruled out for the future. However, many nations will continue in the tradition of viewing the enterprise as much as a community as a commodity.[11] In this tradition, ownership of important chunks of a nation's wealth-producing capacity is seldom left to the vagaries of the market. If foreign shareholdings in a parent company begin to exceed 50%, it is often a sign that the firm is losing its status of parent company and becoming the subsidiary of a foreign corporation.

Whatever the ownership and control of the parent company, at the level of the subsidiaries the relationship is crystal clear. Control and ownership rest firmly with the parent. GM controls and owns 100% of Opel (Germany), Vauxhall (UK), and all its home and overseas subsidiaries. Ford controls and owns 100% of all its overseas subsidiaries except Ford Canada, in which the Canadian government has a minority shareholding. All IBM foreign subsidiaries are wholly owned and controlled by IBM World Trade Corporation in New York, which is wholly owned and controlled by IBM Corporation (also headquartered in New York). European and Japanese multinationals have, by and large, been adopting the same pattern, except where the host government mandates a local partner and/or local shareholdings.

Why do ownership and control matter? Ownership has several important implications. First, the subsidiary's profits accrue to the parent, an entity located in another nation. These profits represent a continuing foreign liability for the host country, and these liabilities increase over time through the reinvestment of earnings. When the profits are remitted to the parent as dividends, usually year after year, the outflow is a substraction from the host nation's GNP and a debit item in its balance of payments. In relation to the initial investment (part of which may have been financed locally through local borrowing), these profits may be substantial. In the famous case of General Motors-Holden Ltd., the Australian wholly owned subsidiary of General Motors, it was found that the 1953-54 profits after taxes amounted to 560% of the original dollar investment and that the dividend declared to the parent, at 46% of profits, represented about 8% of the dollar export receipts in the Australian balance of payments for 1954-55.[12] While this outflow does not constitute a complete picture of the balance of payments effects of foreign investment for the host nation,[13] it is noteworthy that such outflows are a distinctive feature of MNC subsidiaries: locally owned companies do not accrue profits for a foreign parent.

Reproduced with permission of copyright owner. Further reproduction prohibited.

Second, the fact that a multinational group is owned in a home nation means that the company is normally committed to pay dividends in the currency of the home nation and usually presents its consolidated accounts in that currency. This means that it has a home currency. Assets and cash flows labelled in other currencies carry a risk premium compared to those labelled in the home currency.[14] This may affect the firm's behavior. Everything else being equal, the MNC will prefer to keep its liquid assets in its home currency, and will also prefer to locate new facilities in its home nation.

Third, host nation citizens and institutions, should they wish to, cannot participate directly in the ownership of the local subsidiaries of MNCs based elsewhere, since the shares of the subsidiaries will be entirely in the hands of their parents and will not be traded. This deprives host nation interests and the host government of the possibility of exercising an influence, through local shareholders, on the local subsidiaries of the MNCs. Host nation people can buy shares in the parent company, but they are unlikely to have much voice at that level.

The implications of control are no less important. First, although a subsidiary will have, like any local company, a board of directors, the members of the board are selected by, or with the active involvement of, the parent company. The subsidiary's board may contain some of the host nation's most prominent citizens, but their role can only be minor, given that the company is owned and controlled elsewhere. Often the parent company does not expect the local boards to take initiatives, and in many cases half of the board is composed of people from the parent company.

Second, like any local company, a subsidiary will have a CEO and senior executives, but they will have been selected and appointed by the parent company. Where the CEO is not an expatriate from the parent company, he or she will often be a local who has won the confidence of the parent company by, for example, having worked at the headquarters. Often, alongside local managers, expatriates will be holding important positions and, with direct lines of communication to the headquarters, will be making or overseeing important decisions.

Third, even where a subsidiary has been assigned a "global mandate" (i.e., responsibility for a business segment or product line not just in the host nation but throughout the world), certain powers are reserved by the parent and often include: authority to commit capital expenditure above certain limits; borrowing or raising of money; senior appointments; and major new commercial developments.[15]

Fourth, where there are regional headquarters to coordinate and oversee the national subsidiaries in the region, these tend to be headed by home country nationals from the parent company.

Fifth, non-expatriate managers and employees at a subsidiary may have a feeling of second-class citizenship. If a high proportion of a nation's economy is controlled outside, and if this includes strategic or key sectors

Reproduced with permission of copyright owner. Further reproduction prohibited.

(i.e., sectors which are quantitatively important, which affect many other sectors or which are key to the future because of high income elasticity of demand, rapid technological progress, and fast growth in labor productivity),[16] it becomes questionable to what extent the nation can really maintain its integrity, coherence, and political independence. Finally, if one accepts that a nation's most important resource is its human resources and that "the power to create wealth, for an enterprise, is the power of decision,"[17] it follows that a nation cannot develop its potential to the fullest if there are not enough nationally based enterprises in which its citizens can exercise the power to make final and integrated decisions.

Global companies are far from being international, multinational, or transnational when it comes to the locus of their ownership and control. If they are viewed by host nations as foreign entities rather than as manifestations of a new world order that transcends national divisions, it is not without justification. It is perhaps for this reason that global enterprises stress "good citizenship" in host nations. However, good citizenship seldom goes so far as the home base giving up any of its control or ownership.

The People—A corporation is not only its assets and operations, but also its people. For a global enterprise to have "no country to which it owes more loyalty than any other," its management and workforce must be multinational. What is the reality?

Except for companies such as Nestlé and Shell, the majority of the total number of employees in the corporation is employed in the group's home nation and are home nation citizens. When it comes to positions in senior management or on the parent company's board of directors, the percentage of foreigners (non-home-nation citizens) is not only significantly lower than the percentage of foreigners in the total number of employees, it is minuscule. GM and Ford do not have foreigners on their boards. IBM has only two foreign nationals (one Swiss and one German) sitting on its main board. Only two non-Americans have made it to the top of IBM's managerial hierarchy and they have become famous because of that.

Due to the close geographical proximity of European nations, the EEC's integration, and the fact that educated people in Europe speak a number of languages, European-based global companies tend to be more internationalized in their senior posts. German companies, which have a dual board structure, often have one Swiss and one Dutch national as members of their supervisory boards. Nestlé has an all-Swiss board but a number of non-Swiss senior managers. ICI has one German, one Japanese, and two American (non-executive) directors, and claims that 40% of the top 170 executives are not British. Even so, this still means that ICI's home nation (the UK), with 40% of the total number of employees, fills 60% of the top posts.

The national character of top management is even more pronounced in Japanese companies. Sony is the only major Japanese manufacturer with

Reproduced with permission of copyright owner. Further reproduction prohibited.

foreigners on the main board. Local managers in Japanese subsidiaries have a much more "subsidiary" role than those in other MNCs. In Japanese subsidiaries in the U.S., the CEO is often Japanese, the American managers hit the promotion ceiling quite soon and are excluded from the inner counsels, and the senior American managers rarely call the shots and are frequently relegated to playing a high-profile public-relations role.[18] There is nothing surprising in all this, since he who pays the piper calls the tune. The global enterprises of today are usually national or binational at best, especially at the senior management level. The few foreigners admitted to the parent board or top posts are admitted precisely because of their compatibility with the national style and character so that their presence should have little disruptive impact on the national character of the firm.

Legal Nationality and Taxation—In legal terminology, there is no such thing as a multinational or global company. At present, there is no international law under which a transnational or supranational company can be formed and have legal existence in several nation-states. Even the European Community, arguably the most integrated group of nations in the world, has so far, after many years of discussion, failed to agree on the legal basis for a European company. Companies can only be formed under national law, and they acquire the nationality, citizenship, or domicile of the country under whose law they are incorporated. The parent company has the nationality of the home nation, the subsidiaries of the respective host nations in which they were created. Each is subject to a different national law, which determines the legal limitations on its behavior. Thus legal nationality affects corporate behavior, since what is legal in one nation may be illegal in another (e.g., a company's buy-back of its own shares or the creation of "floating charges" on its assets as collateral for borrowing); and what is mandated in one nation (e.g., codetermination in German law) may not be required in another.

The separate legal personality of the parent and its subsidiaries means that the parent is not automatically held liable for its subsidiary's liabilities. The concept of limited liability applies to any shareholder, whether that refers to a private individual or a parent corporation. This means that the global enterprise is able to hide behind the legal principles of separate legal personality and limited liability to avoid taking responsibility for the actions of a subsidiary that it owns and controls. In a limited number of cases (such as the Bhopal disaster), the courts have been able to "lift the veil" and to hold the parent or the group responsible. In general, however, this is exceptional and the law on corporate groups is not well-developed.[19]

Under international law, legal nationality is a necessary condition for obtaining the diplomatic protection of the State concerned; however, it may not be a sufficient condition.[20] In order to rule out tax havens and flags of convenience, international law seems to require, in addition to legal nation-

Reproduced with permission of copyright owner. Further reproduction prohibited.

ality (which is simply a matter of place of incorporation), a real and effective link between the corporation and the State. Evidence of such a link can be constituted by control or ownership.

The case of MNCs raises an interesting question for which there is no clear legal answer. If, for example, IBM Germany needed diplomatic protection in Eastern Europe, to whom should IBM turn to for such protection: the U.S. Government or the German Government? Much would depend on the circumstances. The company would probably prefer protection by the more powerful nation's government. MNCs from small nations may prefer U.S. diplomatic protection if they have significant operations in the United States. It is significant that in the case of exports of automobiles from Japanese "transplants" in the U.S. to the EEC, the Japanese have apparently left the matter of their diplomatic representation in the EEC to the U.S. Government.

Another interesting question arises in cases where a company changes its legal nationality for tax or political reasons (usually by creating, in a country with a favorable tax and/or political regime, a holding company which then owns the former parent company and/or its subsidiaries—as was the case with the "moves" by Tetra Pak to Switzerland, by Jardine Matheson to Bermuda, and by the Hong Kong Bank to Britain). To which government will it turn to for diplomatic protection in case of need? The government of the country where its holding company is now incorporated (and where its head office may, but need not, be located) or the government of the former country where the bulk of its assets, operations, and people are still located? In the case of moves from Hong Kong to Britain or Bermuda, the problem is masked by the fact that the sovereign power remains the same (Hong Kong and Bermuda being British colonies). Tetra Pak, cited by Ohmae as an example of the freedom to move, is a company which achieved fiscal emigration from Sweden to Switzerland. It is now facing a very large fine (US$85 million) by the EEC Commission for anti-competitive behavior.[23] Will it get diplomatic protection from the Swedish or from the Swiss government, or neither?

The extra-territorial application of U.S. laws in the territorial jurisdiction of other sovereign nations results partly from the problems surrounding the definition of corporate nationality. For the purposes of certain laws, the U.S. has adopted the position that the law should apply to all U.S. nationals, and U.S. nationals have been defined to include not only U.S. corporations but also their subsidiaries abroad. This has led to conflicts with nations which define corporate nationality in terms of the place of incorporation. Where a clear conflict of laws can be demonstrated, the host nation has been able to intervene and to prevail by invoking the principle of territoriality. It should be noted, however, that, in the absence of such conflict of laws, there is little that the host nation can do about the fact that instructions and decisions flow one way in any hierarchical system.

Reproduced with permission of copyright owner. Further reproduction prohibited.

Taxation is a key interface between the State and the business corporation. From the point of view of a nation's tax authorities, there are key differences between a home-based company with foreign operations and a company which is a subsidiary of a foreign parent corporation. In nations that adopt the worldwide principle of taxation (the U.S., Britain, and most OECD countries), the home government can *de jure* tax the home-based MNC on its worldwide earnings by virtue of the fact that the subsidiaries' earnings accrue to the parent. Tax deferral (whereby foreign subsidiaries' earnings are not taxed until remitted to the parent as dividends) and foreign tax credits (whereby tax paid abroad can be credited against home country tax liabilities) are concessions with regard to the principle of worldwide taxation. With regard to a foreign-based MNC, the tax authorities can, as a rule, only tax the locally generated earnings of the local subsidiary, since the parent and sister companies are not owned by the local company and hence their earnings do not belong to it. *De facto*, there are also important differences. The home government has access to the consolidated accounts of the home-based MNC, can require additional data through the head office, and can tax the group accordingly. With the local subsidiary of a foreign-based MNC, the problem of determining what the local earnings are is clouded by the problem of inter-country transfer pricing[22] and various charges (royalties, license fees, etc.) payable to the parent and sister companies. Consolidated accounts are of little help since there is no right to tax the worldwide earnings of the group, and access to the head office is more distant. With a home-based group, transfer pricing may alter the inter-country allocation of declared profits, but as long as there is no leakage out of the system, it does not alter the group's total profits and hence its potential tax liability to the home government.

In sum, with home-based MNCs, not only is the tax potential higher, the scope for tax avoidance or evasion is smaller. It is important to note that most advanced nations have tax rules or Treasury regulations designed to prevent fiscal emigration, the transfer of assets to foreign entities, and other moves where gaining a tax advantage is the main objective, with unauthorized moves resulting in severe tax liabilities on hidden reserves or deemed gains and/or criminal penalties.

Although the global firm is exposed to many jurisdictions, it usually has a home government and a home tax authority. It therefore has a legal and a fiscal nationality that matters to it more than others.

The International Competitive Advantage of the Firm

Consider the German chemical industry, the Swiss pharmaceutical industry, and the Japanese computer industry, which are world-class leaders in their fields. Assume that the American chemical industry, or pharmaceutical industry, or computer industry is in danger of losing its international com-

Reproduced with permission of copyright owner. Further reproduction prohibited.

petitive advantage because the environment in the home nation has become unfavorable. Couldn't a U.S.-based company, by establishing a local presence in Germany, Japan, or Switzerland, tap into the foreign nation's competitive advantage and so offset its own, home-based disadvantage? The problem lies in the relativity of competitive advantage. Although tapping into German, Japanese, or Swiss innovations, ideas, skills, aptitudes, attitudes, dynamism, or other sources of advantage would probably add to the U.S.-based firm's advantage, it is likely that the German, Japanese, or Swiss companies would benefit even more from these sources of advantage. They are, after all, operating in their own environment and system. They have been there for a long time and are expected to remain there for a long time. The companies are staffed, managed, and owned by citizens who are members of the same national community. Their self-respect, as well as the esteem in which they hold each other, is intimately connected with this membership, and loyalty to the nation is a paramount commitment of all citizens.[23] Social memory, as a system for ensuring serial equity among groups by remembering who made sacrifices in the past,[24] works only if there is long-term (if not permanent) membership and involvement; it works better if there is a multiplicity of ties and networks. Social memory functions in a nation but not between nations. Foreigners and foreign entities may never gain full acceptance into a national community and its social memory—or even if they do, it requires a very long time. Moreover, top people like to deal with top people, and the local head of a foreign-based group may be perceived as a subordinate who has to report elsewhere. Thus, irrespective of whether explicit "discrimination" exists or not, it is unlikely that a U.S.-based firm can, even with a local presence, plug into a foreign nation's sources of advantage as effectively as competitors who are based, owned, controlled, and managed locally. Furthermore, competitive advantages that can be derived from locating in third countries (e.g., low cost locations) are equally available to the German, Japanese, or Swiss companies; they cannot, therefore, represent a source of *differential* advantage for American firms. Thus: *The primary source of a company's international competitive advantage lies in its home nation; foreign sources of advantage can supplement national sources but cannot be sufficient as a substitute.*

The possibility of joint ventures or strategic alliances with world-class competitors does not alter the thrust of this argument. To get a possible partner interested in cooperating with one, one must have something to offer, i.e., one must possess a competitive advantage. In industries characterized by rapid technological progress, the nature of this advantage will be intimately linked to continuous innovation, improvement, and investment; and hence it is likely that competitive advantage will be generated in the home nation. In sectors with slow technological developments, competitive assets such as distribution networks or market shares erode less quickly with time; and hence they may be more independent of the current status of

Reproduced with permission of copyright owner. Further reproduction prohibited.

the firm's home base. Thus: *In industries with rapid technological progress, strategic alliances are no substitute for creating and sustaining competitive advantage at home.*

What about companies like Nestlé which has 95% of its total operations located outside the home nation or Philips with an estimated 85% of assets abroad? These companies may be owned, controlled, and managed from the home nation, but surely the home base, with such a small weight in total operations, is not central to the maintenance of competitive advantage? The competitive advantage of Nestlé may be based partly on its national qualities, real or perceived: Swiss quality, Swiss standards, Swiss thoroughness or meticulousness, Swiss management, and the reputation of all things Swiss in the eyes of foreigners. If so, severing the links with the home nation and with Swiss people and attitudes would severely disadvantage the company. The impact of the home nation's reputation is likely to be more pronounced in some service industries than in manufacturing industries. In buying services, which are "experience" goods rather than "inspection" goods, reputation can be of decisive importance, and reputation is often national. In banking, moreover, the home nation's performance affects its banks' ability to raise funds in international markets and the price they have to pay. Thus: *The reputation and qualities of the home nation may be an important source of advantage for a firm operating internationally, in spite of the fact that the home nation may account for only a small percentage of worldwide assets and operations.*

It is noteworthy that almost all of the so-called MNCs (for which such data is available) do the bulk of their R&D in the home nation. For example, a 1983 survey of 23 German MNCs showed that 83% of their R&D personnel was concentrated in the home nation (which accounted for only 65% of total employment).[25] Du Pont has 90% of its R&D personnel in the U.S. in 1989 (compared with 65% of its total assets and 76% of its total employment).[26] The same tendency appears to be more pronounced in small country "multinationals": for Philips, 40% of R&D is said to be concentrated in the home nation, which accounts for only 15% of total assets. Another way of saying this is that, within the MNC, the R&D intensity is higher for the parent company at home than for overseas subsidiaries. Thus, for the sample of German MNCs mentioned above, R&D intensity (in terms of manpower) was 9.6% in Germany compared to 3.6% abroad. This is nothing new. The nation where the bulk of a company's R&D takes place indicates where the strategic core of its innovative effort lies. Innovation requires integrated and strategic decision making. In innovative activities, the home nation is the locomotive, assuming a leading role out of proportion to its weight in total operations. Thus: *The home nation is the center of a firm's innovative efforts, and it is where strategic and integrated decisions are made.*

Consider the case of "multinationals" who suffer from an unfavorable

Reproduced with permission of copyright owner. Further reproduction prohibited.

home environment but who have "strong" overseas operations established at a time when the home nation was more dominant and dynamic. Couldn't the group maintain its overall strength by relying on its overseas strength to compensate for its domestic weakness? The problem again lies in the relativity of competitive advantage. The overseas subsidiaries in Germany or Japan may indeed be strong in relation to the parent company, but this does not mean that they are necessarily strong in comparison with their local (German or Japanese) competitors. What happens then? A senior executive of the Japanese subsidiary of a foreign chemicals company commented: "Twenty years ago I spent every night at the best restaurants in Tokyo, being entertained by the chief executives of major Japanese companies, who wanted to gain access to our technology through joint ventures or licensing. . . . Today . . . no one asks me out to dinner anymore."[27] It becomes only a matter of time before profits and market shares tumble.

It is possible that, at some point in time, an overseas subsidiary may show a better profitability than the parent company. This may be the result of the home and host nations being in different phases of the business cycle. It may be due to the fact that the subsidiary is the global home base for a segment or product line that is going through good times. More commonly, however, it may simply be the consequence of the particular foreign market being more protected that the home market; for example, in the 1980s, the European automotive markets were more protected against Japanese competition than was the U.S. market. Whatever the case, this differential profitability is likely to be a temporary phenomenon and will not contribute to the competitive advantage of the parent or the subsidiary unless the profits are reinvested in continuous improvement and innovation.

Another problem is that when the parent company is under siege, it will milk or sell its overseas subsidiaries to save itself. Faced with falls in earnings in the short term, the likelihood of hostile raids, criticism by shareholders, and a generally short-term mentality, U.S. companies have been selling their stakes in Japan.[28] For example, faced with the need to reduce borrowing and with a quarterly loss, Avon Products sold its Japanese subsidiary for $408 million in February 1990. In 1991, Chrysler sold half of its stake in Mitsubishi Motors for $592 million to smooth over a third-quarter plunge in profits. Honeywell sold half of its stake in Yamatake-Honeywell for $407 million in 1989. And in Britain, the engineering group Davy Corp., faced with a half-year loss and a collapse in its share price, sold its German subsidiary, the jewel in the crown, Zimmer AG, to the Metallgesellschaft group for DM 228 million in cash, in February 1991.[29] Thus: *Lack of international competitiveness at home is unlikely to be compensated by overseas operations, but may result in the harvesting of the latter.*

Finally, what about the possibility of shifting the home base to take advantage of a more favorable environment? If a company has 60% or more of its assets, operations and people in one country, it is hard to see how it

Reproduced with permission of copyright owner. Further reproduction prohibited.

could transfer the 60% wholesale to another country. Shifting the home base is not the same thing as shifting a *divisional* headquarters or changing the registered office (i.e., the legal address) of the holding company for political or tax reasons. What can be located abroad is the home base for a particular segment, business or product line, especially when it is new to the company and/or the group's involvement results from the acquisition of an existing company abroad, which then becomes the home base for that business. Despite its global mandate, however, the subsidiary would still have to report to a parent company. Moreover, the shifting of the registered office for tax avoidance purposes is severely controlled. As for *corporate* headquarters, it is simply not effective to have it far removed from the nation where most of the operations are located. Thus: *The company as a whole cannot easily move its home base; it needs its home nation and has a strong interest in contributing to make it a dynamic and competitive base*.

Implications and Conclusions

Well-known companies such as IBM, General Motors, Sony, Honda, Matsushita, Philips, Unilever, Nestlé, Thomson, Hyundai, Daewoo, Tatung, and so on have been variously called by many names—the international, multinational, transnational, global, or stateless corporation. Are these companies national or are they what these adjectives suggest, something that transcends and is separate from nations? To answer this question, the following criteria should be applied:

- In which nation or nations is the bulk of the corporation's assets and people located?
- By whom are the local subsidiaries owned and controlled, and in which nation is the parent company owned and controlled?
- What is the nationality of the senior positions (executive and board posts) at the parent company, and what is the nationality of the most important decision makers at the subsidiaries in host nations?
- What is the legal nationality of the parent company? To whom would the group as a whole turn to for diplomatic protection and political support in case of need?
- Which is the nation where tax authorities can, if they choose to do so, tax the group on its worldwide earnings rather than merely its local earnings?

The relevance and significance of corporate nationality are revealed as one goes over the questions. These criteria usually produce an unambiguous answer: that it is a national corporation with international operations (i.e., foreign subsidiaries).

There are two exceptions to this answer. First, there are, as noted earlier, the binational companies which are owned, controlled, and staffed in two

Reproduced with permission of copyright owner. Further reproduction prohibited.

home nations, and whose assets have two centers of gravity. Secondly, there are the firms from small nations, for whom the home nation accounts for a small percentage of total assets and operations; however, companies such as Nestlé, Philips, and Ericsson satisfy all the other criteria for being Swiss, Dutch, or Swedish, and it is noteworthy that the Swiss, Dutch, or Swedes have no doubt in their minds as to the national character of these enterprises.

Thus, apart from the binational companies, there are no multinational, transnational, or global enterprises, only national firms with international operations.

Who Is Us?—If we think in terms of MNCs, TNCs, or global firms, the power of associations and connotations will draw our attention towards the imagery of a single world economy, in which Reich's question "Who is us?" becomes irrelevant. In this borderless world, the behavior of the global corporation is said to be driven solely by pure economic rationality on a global scale. If a nation loses out in, say, the global firm's location of new investment, the possibility of bias or discrimination is *ex hypothesi* ruled out, because of this assumption of "rationality." If, on the other hand, we recognize these firms to be national firms with international operations, we will ask whether they are American, Japanese, German, Swiss, Dutch, British, or what? If corporations are no longer transnational or global, then depending on who we are, they are either national or foreign.

What about foreign companies that undertake manufacturing, even R&D and design, in the U.S. and that provide employment to American workers and managers? Reich implies that these firms may be more American than American corporations that have international operations.[30] To address his question "who is us?" we can apply the criteria of corporate nationality and compare Toyota, Nissan, and Honda (which have production facilities in the U.S.) with Ford and GM (which have extensive international operations).

- *Center of Gravity*—In 1989, GM had 76% of its assets and 69% of its employees in the United States. Ford and GM produce 55 to 65% (in units) of their worldwide output of cars in the United States. For Honda, arguably the most internationalized of all Japanese automotive manufacturers, the U.S. represented only 22% of total manufacturing workers worldwide as of March 1990 (63% of total assets and total employees being concentrated at home). Honda, Nissan, and Toyota produce 70% to 90% (in units) of their total output at home. Thus, the U.S. matters much more to U.S. firms than to their foreign rivals; the destiny of U.S. companies is more tied to that of the U.S., and they are likely to be particularly receptive to demands and pressures from the United States.
- *Ownership and Control*—Honda US, Toyota US, and Nissan US are owned and controlled by their parent companies in Japan, which themselves are owned and controlled in Japan. The subsidiaries' profits accrue to the parents. The home currency is the Yen. The U.S. subsidiaries are

Reproduced with permission of copyright owner. Further reproduction prohibited.

not listed on U.S. stock exchanges and so cannot be owned directly by Americans. The important decisions are made by the Japanese. GM and Ford are majority owned by U.S. citizens and institutions, and are controlled from the U.S. by Americans. Their worldwide profits accrue to the United States. The home currency is the dollar.

- *People*—There are no non-Japanese on the parent company boards of Honda, Toyota, and Nissan or in their top management positions. The U.S. workers and managers of these companies are a minority compared to their counterparts in Japan. With Ford and GM, top management and the parent board are American, and the U.S. workforce is the majority.
- *Legal Nationality*—Toyota US is a U.S. corporation because it is incorporated in the U.S., but the parent company is Japanese. In the event of a conflict of interest, Toyota may ask the Japanese government or the Japanese lobby to intervene in the United States. Ford or GM could never call on a foreign government to intercede with the U.S. government.
- *Tax Domicile*—Ford and GM pay U.S. taxes on a worldwide basis; Honda, Toyota, and Nissan pay U.S. taxes on a local basis, i.e., on the basis of the income in the United States.

Thus, because of the special nature of the links between the corporation and its home nation and because of the citizenship of the majority of its owners, managers, and workers, a "national" company with international operations is "one of us." It is entitled to support from its home nation, if only because it usually pays more tax to the home government than to any other government.

Corporate Nationality and National Competitiveness—Companies tend to concentrate their mainstream innovative efforts and strategic decision making in the home nation. This means that the most important and the most skill-intensive jobs will be located in the home nation, and that exports will be generated on the back of innovation. This process enhances the nation's standard of living.

Foreign companies may make the U.S. their home base for a particular business segment or product line, but their home nation remains their home base for the corporate group as a whole. Foreign companies may indeed provide jobs and carry out R&D, engineering, or design in the U.S., but it would be interesting to compare the R&D intensity of foreign-owned subsidiaries in the U.S. with that of their parents and that of U.S. corporations. It is also noteworthy that, even with such a localized manufacturer as Honda in the U.S., recent federal and private studies show that the "domestic content" or local valued-added of its cars is not as high as had hitherto been thought.[31]

Implications for Theory, Research, and Pedagogy—If, as I have argued, the multinational corporation does not exist, how would this affect the theory of the multinational corporation? The existence of the theory could

Reproduced with permission of copyright owner. Further reproduction prohibited.

be preserved by arguing that the expression "MNC" is to be understood as merely a shorthand for national or binational firms with international operations. If so, uniformity will be replaced by diversity. The search for a universal theory will no longer be possible, once the umbrella term has been discarded. Instead, we will recognize that there are differences and similarities in the international operations of firms from different home nations, with regard to:

- Why do they go abroad? What are the differences and similarities in the motives for going abroad, and in the transferable competitive advantage that makes it possible to go abroad successfully?
- How do they operate abroad? Do firms from different nations show different patterns as between exporting, producing abroad, licensing, joint venture, subcontracting, and so on?
- Where do they do? Do firms from different nations have different geographical preferences?

Similarly, the search for universal, uniform recipes on how to manage the transnational or global enterprise would give way to a more complex and variegated view. Not only do parent companies from different home bases display different practices and values, but, within the same corporate group, there may be differences between the parent and each of its overseas subsidiaries, depending on the strength of local institutions, practices, and values.[32] And these differences may continue to exist as long as nations continue to be the focus of people's sense of belonging and identification as well as a major component of their personal identities. This view of the nation as "community" (but as national community, endowed with the attributes of independence, continuance for the future and sovereignty) to which people belong (and to which their forefathers belonged and their descendants will belong) may help to explain why the removal of barriers need not undermine the vitality of nations. This is confirmed by the experience of history, which shows that the world was more borderless in the period 1870-1913 than it is today, in the sense that the barriers to trade, capital movements, and migration were lower then.[33] Yet nations, and the frenzied nationalism that led to war, had not withered away.

Although this article has focussed on the concept of nationality, to recognize the existence of nations is not the same thing as to advocate nationalism. Nor is the sense of belonging to a national community incompatible with membership of both sub-national communities and of the world community. To quote the founding father of the national school of political economy: "As the individual chiefly obtains by means of the nation and in the nation mental culture, power of production, security, and prosperity, so is the civilization of the human race only conceivable and possible by means of the civilization and development of the individual nations."[34]

Reproduced with permission of copyright owner. Further reproduction prohibited.

References

1. R. Reich, "Who Is Us?," *Harvard Business Review* (January/February 1990).
2. K. Ohmae, *The Borderless World* (London: Collins, 1990), p.194.
3. Parentage of the term was first pointed out by P. Baran and P. Sweezy, *Monopoly Capital* (London: Penguin, 1966), p. 192. It was confirmed by D. Fieldhouse, "The Multinational: A Critique of a Concept," in A. Teichova et al., *The Multinational Enterprise in Historical Perspective* (Cambridge: Cambridge University Press, 1986).
4. C.P. Kindleberger, *American Business Abroad* (New Haven, CT: Yale University Press, 1969), lecture 6.
5. The question of the nationality of the MNC was first raised by Stephen Hymer in his doctoral thesis, published posthumously. See S. Hymer, *The International Operations of National Firms* (Cambridge, MA: MIT Press, 1976). See also Y.S. Hu, *The Impact of U.S. Investment in Europe: A Case Study of the Automotive and Computer Industries* (New York, NY: Praeger, 1973), pp. 252-267.
6. R. Reich, *The Work of Nations* (New York, NY: Knopf, 1991).
7. Howard Perlmutter, "The Tortuous Evolution of the Multinational Corporation," *Columbia Journal of World Business* (January/February 1969).
8. Y.S. Hu, op. cit., p. 256. The reasons for this preferences for 100% ownership are also explained, see pp. 257-259.
9. *Business Week* "The Stateless Corporation," May 14, 1990, p. 57.
10. I. Walter and R. Smith, *Investment Banking in Europe* (Oxford: Blackwell, 1989), pp. 77-79.
11. See Michel Albert, *Capitalisme Contre Capitalisme* (Paris: Seuil, 1991).
12. E. Penrose, "Foreign Investment and the Growth of the Firm," *Economic Journal* (June 1956).
13. The balance of payments effects should include the effects on exports, imports, royalties, licence fees, etc. as well as the outflow of dividends, and the effects should in theory be assessed against what would otherwise have happened.
14. Items denominated in foreign currencies incur translation, transaction, and economic exposures to exchange movements. Items denominated in the home currency only incur economic exposure.
15. See ICI's Annual Report for 1989.
16. OECD, *The Industrial Policy of Japan* (Paris: OECD, 1972), p. 15.
17. J.J. Servan-Schreiber, *Le Défi Americain* (Paris: Denoël, 1967), Chapter 4.
18. *Business Week*, December 17, 1990, pp. 50-52.
19. See P. Blumberg, "The American Law of Corporate Group," and D. Muchlinski, "Group Liability and the Multinational Enterprise," papers presented at the University of Warwick Workshop on Corporate Control and Accountability, June 30-July 2, 1991.
20. See, Barcelona Traction, Light and Power Co. Ltd. (Belgium v. Spain), International Court of Justice 1970, and Nottebohm (Liechtenstein v. Guatemala), ICJ 1953 (111) and ICJ 1955 (93).
21. *International Herald Tribune*, July 23, 1991 and July 25, 1991.
22. *Newsweek*, "How Multinational Firms Use Transfer Pricing to Evade At Least $20 Billion in U.S. Taxes," April 15, 1991.
23. H. Morgenthau, *Politics Among Nations* (New York, NY: Knopf, 1978), p. 494.
24. W. Ouchi, *The M-Form Society* (New York, NY: Avon, 1986), various chapters. A social memory consists of a network of civic, business, and government associations/institutions that not only can remember who has been cooperative and who has been selfish, but also has the ability to reward and punish when the time comes. A social memory is what was called a sense of community, a sense of responsibility, a sense of

Reproduced with permission of copyright owner. Further reproduction prohibited.

civic-mindedness. Serial equity is equity which is achieved over a long period of time, so that a group that is disadvantaged today knows that its sacrifices will not be forgotten.

25. M. Wortmann, "Multinationals and the Internationalization of R&D: New Developments in German Companies," *Research Policy* (April 1990).
26. Annual report.
27. *The McKinsey Quarterly* (Winter 1990), p. 24.
28. *Business Week*, March 12, 1990, pp. 20-21.
29. *International Herald Tribune*, February 7, 1991 and February 8, 1991.
30. Reich (1990), op. cit.
31. *Business Week*, "Honda, Is It an American Car?" November 18, 1991.
32. The transfer of management practices between national systems is discussed by Richard Whitley, *Societies, Firms and Markets in East Asia* (forthcoming), Chapter 8.
33. *The Economist*, "The State of the Nation-State," December 22, 1990.
34. Friedrich List, *The National System of Political Economy* (originally published in 1841), Chapter 15.

Reproduced with permission of copyright owner. Further reproduction prohibited.

931
693
69
WHARTON REPROGRAPHICS

Environment and Planning A, 1988, volume 20, pages 761-788

The internationalization of engineering consultancies: problems of breaking into the club

P J Rimmer
Department of Human Geography, Research School of Pacific Studies, The Australian National University, Canberra, PO Box 4, ACT 2601, Australia

Received 11 May 1987; in revised form 20 August 1987

Abstract. Changing global trends in the location of engineering consultancies are examined. This reveals the relative decline of firms headquartered in North America and the rise of European-based counterparts. It raises the question of why multinationals located in Asia, Africa, Latin America, and the Middle East have not broken into this Western club in greater numbers. Of particular interest is the relatively poor showing of consultancies based in Japan, Korea, and the Philippines—countries conspicuously successful in exporting construction contracting.

> "[An engineering consultancy devotes itself] to advising the public on engineering matters or to the designing or supervising the construction of engineering works, and for such purposes occupies and employs [its] own office and staff, and is not directly or indirectly concerned or interested in commercial or manufacturing interests such as would tend to influence [its] exercise of independent professional judgement in the matters upon which [it] advises."(1)

Engineering consultancy services are a leading edge in a new and booming international trade (Riddle and Sours, 1984). They belong to a group offering intermediate services—typified by finance, insurance, and real estate (FIRE) and by a wide-ranging group of free-standing business services—that are *inputs* to the production processes of industries (Stanback, 1979; Stanback et al, 1981; Tucker, 1981). Skills previously embodied in either commodity trade or direct foreign investment are now *unbundled* or *splintered* and exported in the form of producer services—a means of overcoming a protection wall for goods. These intermediate or producer services are conceptually and empirically different from those, characterized by international tourism, which are outputs destined for final demand by households or by other types of consumer. Attention is concentrated here on the relatively neglected exports of 'out-of-house' engineering consultancy services, because there is a growing literature on international trade both in tourism and in exchanges of property rights and contractural claims (FIRE) (see, for example, Noyelle and Stanback, 1984; Seow et al, 1984; Tucker et al, 1983).

Exports of engineering consultative services are personally provided and personally underwritten (that is, they are information based and reliant upon an advisory relationship being established between consultant and client). The 'production process' and delivery of problem-solving services by engineering consultants use tangible inputs (for example, transport, computer time, furniture, etc) but they do not export a tangible commodity (Tucker, 1981). Rather than the physical movement of a commodity, the key attributes are data exchange, documentation, and personal mobility. By convention, payment for a consultancy contract (or agreement) is generally time-related and, unlike employment within a firm, is conditional upon the completion of a particular assignment (Kuecken, 1978). It is usually calculated within an overall budget at 'agreed fixed rates' for each man-month (though cost-plus arrangements may be used where the scope of a project is poorly defined; a fixed lump sum where it is specified in complete detail; and a percentage fee where architectural services are involved) (Rimmer, 1984; 1986). According to the services marketing literature, individual, private company, or public corporation buyers "purchase neither tangible objects nor intangible features; rather, they purchase a bundle of benefits—a product" (including ideas, behaviours, and concepts) (Enis and Roering, 1981, page 1).

---

(1) Memorandum for the purpose of incorporating the Association of Consulting Engineers, London, 13 November 1913.

Interest here is centred on engineering consultancy services which are traded across national borders (though the knowledge of domestic trade is equally inadequate). It has not been an instant phenomenon but has been brought about by the progressive expansion of the firm's management perspective from local to global scale. In the examination of this process of internationalization, attention is focused initially on a general survey of international consultancy services, which highlights that the so-called 'indigenous multinationals', based in the nonsocialist countries of Asia, Africa, Latin America, and the Middle East, have had difficulties in breaking into what has hitherto been an exclusive Euro-American club (section 1). As the past lies heavily on the present conduct of the profession, a brief outline is given of the establishment of European and American practice before the trends that are symptomatic of a new phase in engineering consultancy practice are considered (section 2). A distillation of these trends leads to the recognition and development of a universal scale of internationalization among engineering consultancy firms (section 3).

One is, then, in a position to gauge the progress of 'indigenous multinationals' on this scale. Rather than try and encompass firms across all continents, attention in gauging the efficacy of this scale is concentrated on an examination of international engineering consultancy firms based in three Asian countries—Japan, South Korea, and the Philippines—which have been conspicuously successful in exporting construction-contracting activities (Rimmer and Black, 1984). The relatively poor performance, on the universal scale, of Japanese, Korean, and Filipino engineering consultancy firms is considered in terms of their specific organizational structures, personnel, and particular management structures (section 4). A superleague of international engineering consultants, however, has emerged, and one has to consider the even poorer performance of Japanese, Korean, and Filipino engineering consultancies on this scale (section 5). Conclusions can then be drawn about the problems faced by African, Asian, Latin American, and Middle Eastern enterprises, and Japanese, Korean, and Filipino firms in particular, in penetrating the ranks of international engineering consultants. I start, however, with a general survey of international engineering consultancies.

International survey

A major problem in reviewing international engineering consultancies is the fragmentary nature of postwar evidence [see, for example, Association of Consulting Engineers, 1970; Consulting Engineering Services, 1971; Registry of Scientific and Technical Services, 1973; ESCAP (Economic and Social Commission for Asia and the Pacific), 1977]. Hence, much use is made of annual surveys undertaken, since 1978, by the *Engineering News-Record* (1979-1986). As is summarized in table 1 these surveys of initially the 'top 100', then the 'top 150', and later the 'top 200' engineering consultancies provide evidence of the trends in the collective value of foreign billings derived from (a) prefeasibility studies, (b) design, (c) construction management, and (d) design-construct. In addition, it is shown in the surveys that the top international consultants increased the dollar value of their billings for each consecutive year from US$1.5 billion in 1978 to US$3.8 billion in 1983 (a figure including reimbursables, temporary staff transfers, and construction inspection) (table 1). A decline of 10%, however, occurred in 1984 followed by a slight recovery in 1985. Domestic figures are unavailable for the period 1978–81 but their contribution to the total billings of leading consultants increased from 55% in 1982 to 64% in 1985. Domestic profits also moved ahead of foreign returns in 1984 for the first time.

A prolonged world recession has characterized the period under review but the rankings of the major regional markets have remained steady. The Middle East has consistently provided over one-third of all foreign billings, followed by expanding markets in Asia and North America—a phenomenon which highlights that fast-growing populations had to double their infrastructural capacities every twenty-five or thirty years. Saudi Arabia has attracted the largest proportion of the leading consultants,

Table 1. Trends in international engineering consulting derived from *Engineering News-Record* surveys, 1978–1985 (source: *Engineering News-Record* 13 July 1979; 31 July 1980; 30 July 1981; 29 July 1982; 28 July 1983; 2 August 1984; 1 August 1985; 7 August 1986).

Item	1978	1979	1980	1981	1982	1983	1984	1985
Consultants surveyed								
Number	100	150	150	150	201	200	204	200
Billings (US$ billion)								
Foreign	1.5	2.1	2.6	3.1	3.7	3.8	3.5	3.6
Domestic	na[a]	na	na	na	4.6	4.8	5.3	6.5
Total	na	na	na	na	8.3	8.6	8.8	10.1
Average profits (%)								
Foreign	na	na	na	na	6.8	8.7	7.1	7.0
Domestic	na	na	na	na	6.5	6.9	7.2	na
Regional breakdown (foreign billings US$ million)								
Middle East	na	700	940	1100	1266	1342	1118	974
Africa	na	320	597	704	880	817	782	792
Asia	na	na	408	523	733	823	804	919
Latin America	na	180	253	319	382	374	277	456
Europe	na	na	na	na	357	396	356	358
North America	na	na	na	24	118	99	126	141
Leading national markets (percentage of firms active)								
Saudi Arabia	54.0	66.3	na	72.0	67.0	68.0	88.5	54.5
Indonesia	na	na	na	41.4	49.5	48.0	60.0	na
Egypt	na	na	45.0	na	45.0	45.0	na	na
Malaysia	na	na	na	34.0	40.5	41.0	na	na
Nigeria	na	33.3	42.0	39.0	39.0	39.5	na	na
Iraq	na	na	na	na	38.5	34.5	na	na
Philippines	na	na	na	34.0	34.5	32.5	na	na
Consultant nationality (percentage share of foreign market)								
American	42.0	38.0	34.0	37.0	33.8	31.3	29.9	32.0
European	na	na	na	na	48.1	51.5	48.8	47.0
Canadian	na	6.0	na	na	7.8	7.0	8.3	7.3
Japanese	na	na	na	3.1	3.0	3.3	4.8	6.2
Korean	na	na	na	na	na	na	1.5	1.3
Others	na	na	na	na	7.3	6.9	6.9	6.2
Consultants (billings over US$ 50 million)								
Number	na	na	na	na	15	16	10	11
Top ten (%)	na	na	na	na	19.0	19.5	20.0	na

[a] na means not available.

followed by Indonesia. Elsewhere, the spotlight has flitted from Nigeria and Egypt to Malaysia and the Philippines. The share of foreign billings held by American consultants has declined since 1978 (though they continued to dominate all continental markets with the exception of Africa where they were surpassed by the French and the British in 1983). Although the figures are incomplete it would appear that European and Canadian consultants were initially the main beneficiaries of the American decline. No other group exceeded 5% until the Japanese consultants achieved that feat in 1985. As the collective total of consultants in the top 200 who are based in Africa, Asia, and the Middle East has increased since 1984, the challenge to West European and American dominance in international consulting may, at last, be under way.

A problem of the *Engineering News-Record* survey is that it is still developmental. Statistics of the number of firms with foreign contracts in excess of US$50 million and the share held by the 'top ten' firms have only been available since 1982. Nevertheless, the survey does provide the names and addresses of the 'top 200' international consultants. When the cities with two or more representatives in 1984 are mapped, as in figure 1, this information underlines the propensity of international engineering consultancies to concentrate in major nodal centres. The dominance of American and West European cities is evident, with London, Paris, and New York figuring prominently—the distribution in the United States reinforcing the argument that the 'snow belt' cities have been prominent in producer services (Noyelle and Stanback, 1984; Sassen-Koob, 1986; Stanback and Noyelle, 1982). The only counterpole is in East Asia where there is a string of centres around Tokyo. Reshuffling, in the importance of cities on this score, has been occurring but worthwhile analysis is precluded by the survey's failure to provide full addresses of individual firms prior to 1982.

Another major drawback of this developmental survey for comparative analysis has been that the base has varied from 100 firms in 1978, through 150 firms in 1979, 1980, and 1981, to 200 firms in 1982, 1983, 1984, and 1985. Hence, for the purposes of making comparisons between countries over time, attention is concentrated on the 'top 100' firms to maintain comparability over the period. As is shown in table 2, the dominant position of the US firms among the 'top 100' firms has been progressively whittled away since 1978. This has been

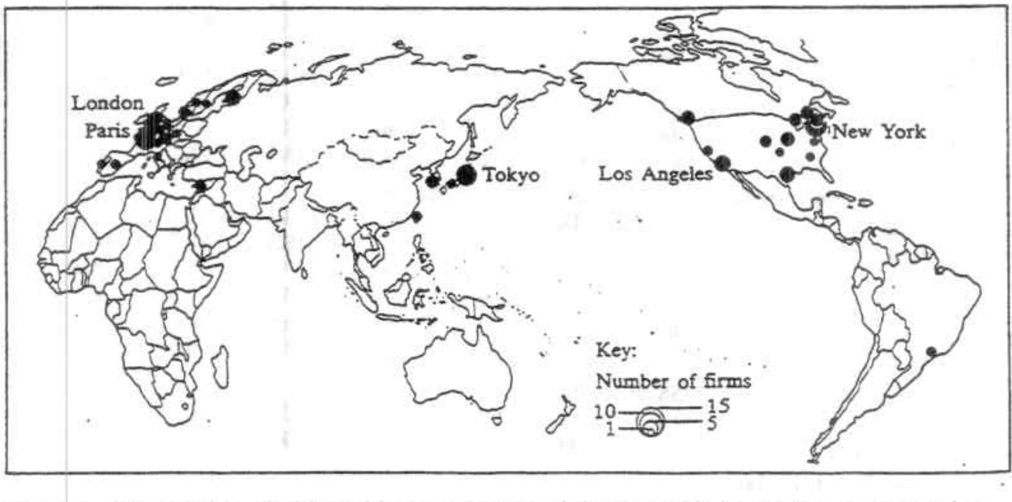


Figure 1. Distribution of cities with two or more of the 'top 200' international engineering consultancies, 1984 (source: *Engineering News-Record*, 1 August 1985).

attributed to the fact that they have lost their traditional edge in high-technology areas though, in part, the decrease is also because of greater participation of non-US-based firms in the survey. As noted, the main beneficiaries have not been, contrary to the case of construction contracting, firms from the Middle East, Latin American and, above all, Asian countries. This evidence suggests that these countries, despite their apparent comparative low-cost advantage in the supply of labour-intensive services, depend largely on the supply of producer services from North America and Western Europe (Tucker, 1981).

European firms, notably those from the United Kingdom and France, concentrating on key energy and manufacturing sectors, have benefitted most from the decline of the US firms. These British and French firms, unlike their US counterparts, have not had a large domestic market and have been forced to tap export markets (Civil Engineering Economic Development Committee, 1978). This is borne out for the case of the United Kingdom, if one draws upon fresh material on overseas regional offices, maintained by her international consultants, to supplement the *Engineering News-Record* surveys. When this material is mapped, as in figure 2, it emphasizes that the consultants from the United Kingdom, like their French counterparts, have capitalized on their economic affiliations with former colonies—Australia, Nigeria, East Africa, Southeast Asia, and a host of Middle East countries (Guttman, 1976). The enigma, given the United Kingdom's strong showing, is the relative failure of Japanese firms to make an equivalent impact, especially as the country has developed throughout the world powerful marketing operations for commodities. Although there are signs that Japanese consultants are, at last, making an impression in the 'top 200' rankings, their leading organizations (and their Korean counterparts) will have to move from the traditional architectural-engineering areas and engage in Construction Management (CM)—the leading edge of engineering consultancy services.

Construction Management has evolved from the search for new approaches to combat construction costs, inflation, and the growing complexity of buildings.

Table 2. Distribution of the 'top 100' international engineering consultants, 1978–1985 (source: *Engineering News-Record* 13 July 1979; 31 July 1980; 30 July 1981; 29 July 1982; 28 July 1983; 2 August 1984; 1 August 1985; 7 August 1986).

Area	1978	1979	1980	1981	1982	1983	1984	1985
Canada	8	6	8	7	7	6	7	7
United States	49	44	34	33	29	24	24	21
North America	57	50	42	40	36	30	31	28
France	10	8	11	11	11	10	9	7
Netherlands	3	4	5	4	6	6	7	6
West Germany	5	6	5	7	8	9	9	9
United Kingdom	8	14	17	14	16	18	18	17
Other	9	12	13	14	12	16	15	16
Europe	35	44	51	50	54	59	56	55
Australasia	1	0	0	1	1	1	1	1
Middle East	2	1	2	3	4	3	4	5
Japan	2	3	3	3	2	2	4	8
Korea	0	0	0	1	1	2	2	2
Other	2	1	0	1	1	1	1	0
Asia	4	4	3	5	4	5	7	10
Latin America	1	1	2	1	1	2	1	1

Although there is no universal agreement of a definition of CM it has become the generic term for a broad range of methods. A basic definition which has received widespread acceptance is that CM is

> "an alternative to the traditional Design/Bid/Build and Design/Build methods of project delivery. Construction Management places primary emphasis on integrating the design and construction phases of a project. It is a Team approach from project inception that strives for project delivery within the optimum time, for the most economical cost and with the required quality. The Construction Management method centers on utilization of a Construction Manager, who coordinates and manages the building process and is a member of the Construction Team with Owner/Architect/Engineers and other consultants as the project may desire" (Lindstrom, 1982, pages 24–25).

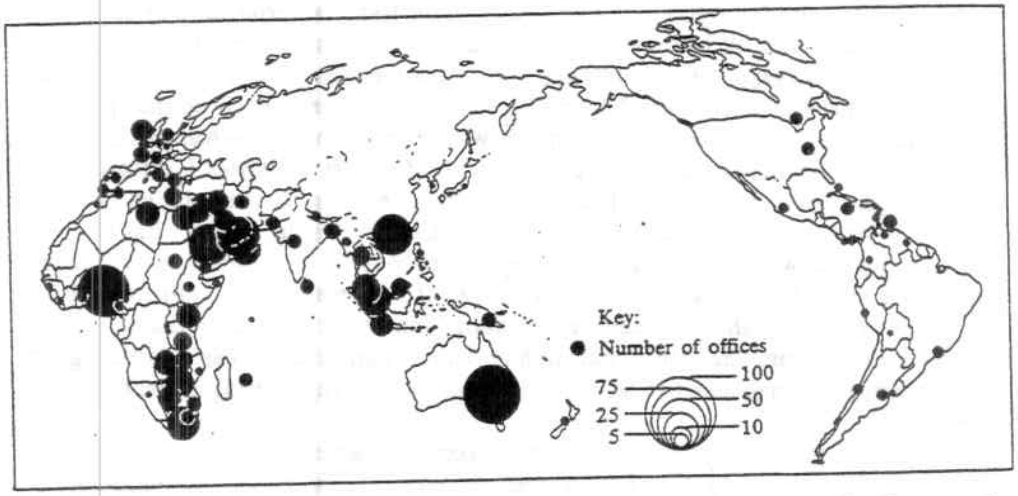


Figure 2. Distribution of the regional offices of engineering consultants with headquarters in the United Kingdom (source: Consulting Engineer, 1985).

Table 3. Distribution of the 'top 50' international construction management consultants, 1978–1981 (source: *Engineering News-Record* 13 July 1979; 31 July 1980; 30 July 1981; 29 July 1982).

Area	1978	1979	1980	1981
Canada	7	5	6	6
United States	27	19	14	15
North America	34	26	19	21
Central and South America	1	1	2	1
France	5	5	7	3
United Kingdom	1	3	5	6
West Germany	1	3	3	5
Other	5	8	11	9
Europe	12	19	26	23
Middle East	1	1	2	1
Other	0	1	0	1
Asia	1	3	1	1
Australia	1	0	0	1
Total	50	50	50	50

CM can be undertaken by either consultants or contractors; I am concerned here with the first type (Pinnell, 1980).

An examination of the 'top 50' CM consultants, published by the *Engineering News-Record* between 1978 and 1981 (table 3), shows that Japan has token representation; Korea and the Philippines have none. It has been another case of Europeans taking over from the Americans. Although the developmental nature of the surveys in the *Engineering News-Record* has precluded any follow-up of these changes over the past four years, they suggest that as the Japanese, Korean, and Filipino engineering consultancies climb one ladder there is another one waiting which may have been pulled up by the time they arrive. Before firm conclusions are drawn on this matter, however, one needs to examine the roots of European and American engineering consultancy practice and the nature of the changes that occurred in the 1970s, as they have done much to shape the contemporary scene.

From Euro-American to global practice

The internationalization of engineering consultancies is not a new phenomenon. An analysis of Bell's (1975) *Bibliographical Index of British Engineering in the 19th Century*, based on 3678 obituary notices in over forty contemporary journals, showed that this experience had been carried overseas in at least 545 instances—over 222 going to India alone (Rimmer, 1980, page 52). A major problem, however, was to separate contracting and consultancy work as the term 'consulting engineer' was undefined during the nineteenth century. Indeed, the distinction was not spelt out in the United Kingdom until the early 1900s when meetings were held to explain the term 'consulting engineer', to prepare rules, and to draw up a scale of fees. The ultimate outcome was the formation of the British Association of Consulting Engineers (ACE) in 1913 with the express intention of embracing qualified engineers who were able to offer clients independent professional judgement for a fee (Consulting Engineer, 1985). It was not, however, a British preserve, as the Germans, if not the French, were strong competitors. Simultaneous to the establishment of the British association in 1913, a group of consulting engineers established European practice by founding the Fédération Internationale des Ingénieurs Conseils (FIDIC)—a body then domiciled in Zurich and representing member associations of ten European countries, including Great Britain.

A parallel professionalization had occurred among engineers and architects in North America, between the Civil War and the First World War, giving rise to work in South America, the Philippines, and even in Russia during the 1930s (Fucik, 1964). As outlined by Levy (1980), urbanization and industrialization in North America had created this demand for a highly educated and trained work force. In turn, they sought the acceptance by clients and by the public of a code of ethics as a means of imposing a new structure on the market, which would guarantee them a secure economic and social position. The oldest and most senior of the associations in the United States was the American Institute of Consulting Engineers (AICE) founded in 1910. This largely New York-based body was matched by state associations. These state associations did not come together as a national body until the establishment of the Consulting Engineering Associations in 1956. Their code of ethics, however, would have been acceptable to the FIDIC (OEEC, 1958). Since their inception, therefore, engineering consultancies have had little to sell except "service, time and knowledge" (Stanley, 1961, page 6).

Invariably, the typical European and American firm which is engaged in civil engineering and architecture began as an individual proprietorship or as a partnership with strong technical leadership from the principals (Logcher and Levitt, 1979, page 7). Although this structure survived the Second World War, European firms had lost the edge to their American counterparts (OEEC, 1958; Morgan and Kamil, 1982). In particular, the distinctive style of American expertise in engineering, fashioned in trying to extricate the motor car from the mud of rural America, became popular after the Second World War. The Marshall Plan and the establishment of the International Bank for Reconstruction and Development (the World Bank) gave great impetus to the translation of this experience overseas. This stimulus was reinforced: by decolonization; by the decision of the US government and of other 'free world' countries to build up infrastructural stock of newly-independent nations; by the establishment of a large group of project-financing organizations, such as the Agency of International Development (AID) and the Inter-American Development Bank (IADB or BID); and by the opportunities for private commercial development in the 'richer' developing countries. The balance of this work tilted towards US consultants because of AID requiring the employment of US consultants on their projects overseas (except where qualified firms existed in the host country). America's active military role in East and Southeast Asia also led to the design of airfields, army installations, and naval bases in Japan, Korea, the Philippines, Thailand, and Vietnam.

During the 1960s American civil engineers "evolved from relatively informal 'fiefdoms' of the partners into internally specialized corporate entities" with functional branches (for example, structural engineering, service, report etc) (Logcher and Levitt, 1979, page 8). Initially, an international division was added to the branch structure in order to control this activity, but in the larger firms it gave way to president–president relationships in a fully-fledged multinational operation—a situation prompted by the increasing geographical remoteness of regional offices and by changes in project size and complexity (Logcher and Levitt, 1979, page 10).

The stranglehold of American companies over international business came to an end in the late 1960s when European consultants returned to the international arena. Recovered from the War, European countries were anxious to shrug off the derogatory 'American pensioner' tag by dispensing with financial aid and technical assistance from the United States (Armytage, 1976, page 306). Hence, European countries bolstered national consultants, through aid projects and direct subsidies, because they were seen as 'stalking-horses' whose legitimate activities provided a cover for the collection of intelligence on upcoming business; this information was vital for national construction contractors. (According to Clarke, 1984, European firms were more likely to develop mother–daughter relationships rather than the miniature-replica organizational structures of their American counterparts.) As European competition undermined the American balance of payments on programs of foreign trade, consultancy services were incorporated in the guarantee program of the Export–Import Bank in 1963, and attempts were made to improve the country's listening posts (US Department of Commerce, 1970). In September 1962, the Secretary of State, Dean Rusk, wrote to all US foreign service posts.

> "I am requesting that your principal aides be alert to, and seek out export opportunities for American business; that you develop the necessary contacts in order to be 'in the know'; that your Mission shows foresight in advising US business not only of impending bids, but also on construction and development projects long before they reach a definitive stage ..." (Kollar, 1963, page 58).

American consultants were at pains to emphasize that their quotations were higher than their European counterparts because they did more work on drawing and design (though, unlike the Europeans, they did not prepare detailed bills of quantities). They also impressed upon the US government that it needed to use its good offices to persuade developing countries to drop the practice of price bidding which offended their American codes of conduct. A program of education was recommended to bring these countries into line with United States (and European) practice on this score (Fucik, 1964). By the late 1960s, however, the state of the art was no longer American or European. It was global.

Symptomatic of this change was the battle over the acceptance of the comprehensive bid, 'package job', or 'turnkey bid' as it is called, which covers the contract price, the design, the construction, and the supervision of the project. This arrangement was actively discouraged by the American Society of Civil Engineers. Some members of the Society were violently opposed to the notion of design-construct because it violated their code of ethics which set fees by negotiation and not by bidding. Indeed, they likened the compilation of the guide to turnkey contracts as "writing a procedure on how to obtain the best prostitute" (Narver, 1975, page 236). Despite their opposition, the turnkey project grew rapidly. Significantly, this opposition was maintained until 1970. Although the American Society of Civil Engineers has been at pains to emphasize that the procurement of engineering services involves other factors in addition to fees, the prohibition or limitation of fee quotations for engineering services has been removed from the code of ethics (Narver, 1975). This change prompts one to look at the nature of international consulting and the effects of competition from the new breed of multinationals, based in Africa, Asia, and Central and South America, and the stranglehold exerted by US and West European firms.

The switch to a corporate focus in the early 1970s has been marked by a greater geographical spread of activities to encompass the whole world. These changes have been accompanied by the evolution of new management structures: (a) multidivisional and (b) matrix. Multidivisional structures were based on areal divisions of the market (or worldwide projects)—a reflection of the growth in the size of projects and in the number of in-house technical disciplines (figure 3). The divisions were based on regional offices in one or more countries (Logcher and Levitt, 1979, page 8; compare Locke, 1984). Corporate headquarters were in the founding office which characteristically contained the top-line officers of the firm and two or more primary support groups (comprising company-wide administrative support services, including personnel, accounting, and data processing, and the marketing or business development function).

As shown in figure 3, the parallel matrix or project operational structure is designed so that:

> "Each [national] office, as a production centre, is typically broken down into disciplinary specialists, who are assigned to one or more projects at any given time. Each disciplinary group is headed up by a discipline manager, responsible for technical oversight and manpower planning and utilization within his disciplinary group. Each project is headed up by a project manager (who may also be a discipline head) responsible for client and interdisciplinary interface planning and scheduling, and overall project costs. It is quite common for project work to be shared between [national] offices, with another [country] supplying the work for a given discipline" (Logcher and Levitt, 1979, page 9).

As a result of this new development, four 'pure' types of international engineering consulting firms can be recognized: international division, miniature replica (or mother-daughter), area divisions, and a matrix structure (see Cardinal, 1980).

It is difficult to assess the performance of these different types of 'pure' firms, as data have not been forthcoming on the pivotal positions occupied by the discipline head and the project manager. This is a handicap, as the growth in project size and complexity and the increasing remoteness of national offices have resulted in the firms functioning almost as independent organizations. As one is not able to identify profit centres, the level of observation is lifted from the microlevel to the macrolevel. At this level the new forms of organization and control, personnel arrangements, and management functions adopted by global engineering consultants can be incorporated in a universal scale of internationalization.

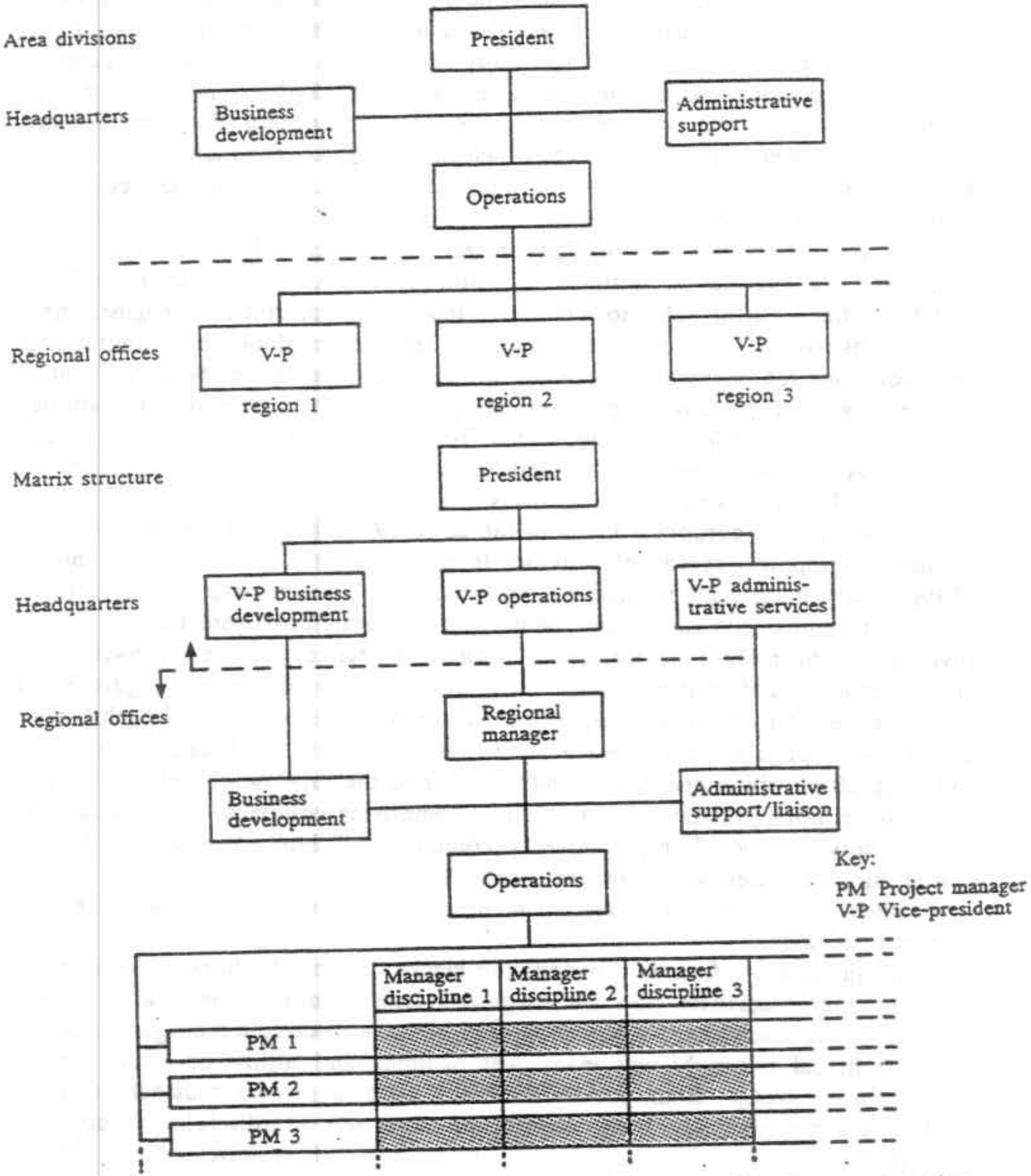

Area divisions
President
Headquarters
Business development
Administrative support
Operations
Regional offices
V-P
region 1
V-P
region 2
V-P
region 3
Matrix structure
President
Headquarters
V-P business development
V-P operations
V-P administrative services
Regional offices
Regional manager
Business development
Administrative support/liaison
Operations
Key:
PM Project manager
V-P Vice-president
Manager discipline 1
Manager discipline 2
Manager discipline 3
PM 1
PM 2
PM 3

Figure 3. The forms of organization and control exercised by engineering consultancies engaged in global practice (source: based on Logcher and Levitt, 1979, page 9).

Universal scale

An examination of engineering consultancies in terms of national 'management cultures' as a reflection of distinctive societies (for example, Europe, United States, and Japan) is too cumbersome. Attention is, therefore, focused on regarding organizational differences in terms of the degree of progress on a universal scale of internationalization. Fortunately, Kobayashi (1981) has developed a conceptual framework to gauge the degree of progress towards internationalization.

It is based on the expansion of a firm's managerial organization from a local to a global scale.

As shown in figure 4, Kobayashi's (1981) conceptual framework has recognized a series of discrete developmental levels which are based on the extent to which a firm's overseas operations are integrated into its overall organization. (To draw on Cardinal's 1980 work, key organizational structures are also attached to each level though all firms do not fit neatly into this scheme as hybrids exist which combine different structures.)

(1) Management by head office (that is, international division), represented by black circles;

(2) Delegation of management to overseas subsidiaries (that is, miniature replica or mother-daughter relationships), represented by grey circles;

(3) Regional coordination of overseas operating units (that is, area divisions involving ties between overseas units and other firms);

(4) Management from a global perspective as it expands beyond a major regional boundary (that is, matrix structure);

(5) Management which combines a global perspective with integration of the activities of overseas operating units into the logistics of the firm as a whole (that is, fully-developed matrix structure).

Having developed this broad conceptual framework Kobayashi (1981) proceeded to specify its key elements.

Eight specific items have been identified by Kobayashi and grouped into three main categories—control, personnel, and particular management functions—to provide a checklist for the determination of the degree to which engineering firms have

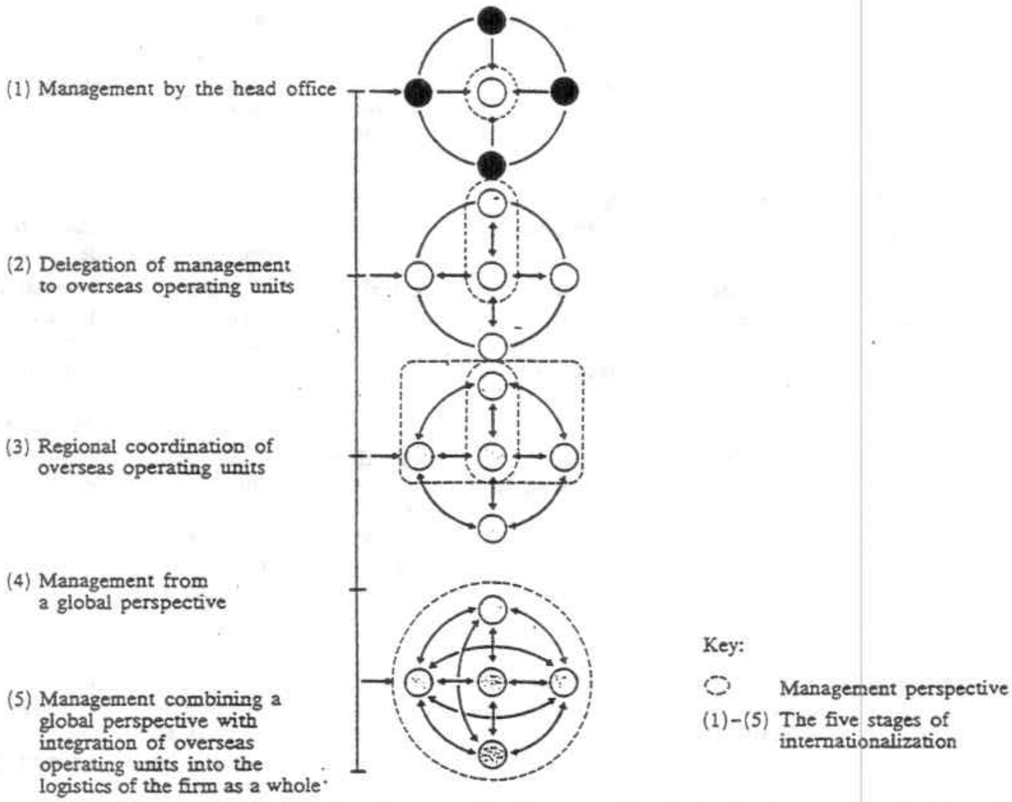

(1) Management by the head office
(2) Delegation of management to overseas operating units
(3) Regional coordination of overseas operating units
(4) Management from a global perspective
(5) Management combining a global perspective with integration of overseas operating units into the logistics of the firm as a whole
Key:
Management perspective
(1)-(5) The five stages of internationalization

Figure 4. Kobayashi's five major stages or developmental levels in the internationalization of management (source: Kobayashi, 1981, pages 12-13).

Table 4. Kobayashi's scheme for determining different levels of internationalization [source: derived from Kobayashi (1981, pages 11–12)].

	Level of internationalization				
	1 home office management	2 overseas operating unit	3 regional coordination	4 global view	5 integrated global
Control					
Planning and decision-making	individual overseas project	regular overseas projects	overseas division	key feature of plan	fully incorporated
Organization control	home office management	creation overseas project	full-time director	separate operational HQ/special subsidiaries	single, unified organization
Personnel					
Overseas training of home-based employees	no consideration	recruit from overseas firms	no systematic program (rotating management)	restrict to overseas project department	fully-integrated
Local managerial employers	no local employees	some local employees	exchange of local managers	higher duties (for example, regional HQ)	managers in home office
Management functions					
Purchasing	one-sided relationship	bilateral relationship	independent	third firm involved	complex network
Marketing decisions	head office	local management	formulate local strategies with head office	overseas office develops original strategy	global strategy
Research and development	head office	local management	coordinated with head office	exchanges between overseas units	head office makes final decision
Financial control	individual project	overseas operating considered separately	regional control	overseas projects in single unified system	overseas and domestic projects controlled as part of single unified system
Fund raising	individual project	overseas operating unit	head office funds from domestic or local capital markets	single system using funds from domestic or overseas markets; domestic treated separately	single system for all projects and capital sources
Investment	no system; case-by-case basis	overseas operating unit	regional control centre	separate investment overseas and domestic funds managed by head office	uniform—no domestic/overseas distinction

been internationalized (table 4). On these criteria, major global European-based and US-based firms, such as Norconsult (Norway) and Louis Berger International Inc (United States), have passed through Kobayashi's five phases. Hence, interest is centred on leading Filipino, Korean, and Japanese firms to gauge the degree to which they have been internationalized. A broad overview is given of engineering consultancies in each country, before attention is focused on specific case studies in order to gauge the efficacy of Kobayashi's universal scale. In turn, these studies provide the basis for an examination of the performance of an individual country's engineering consultancy services. Specific attention is devoted to the enigmatic performance of Japan's engineering consultancies. Before this matter is tackled, however, interest is centred on firms based on the Philippines and South Korea and on their prospects of lessening their dependence on international engineering consultants from advanced capitalist countries.

The Philippines

No Philippine company has been ranked among the top international consulting firms in the annual *Engineering News-Record* surveys between 1978 and 1985. This is not surprising as all the consulting firms are of comparatively recent origin. Prior to Martial Law (1972) the four established consulting firms were engaged mostly on projects for the US Armed Forces (represented locally by the Officer-in-Charge for Construction of the US Navy Corps of Engineers); there was also some government work which occupied local firms established in the 1950s (Redecon, 1981; Kaigai konsarutingu kigyo kyokai, 1982). By 1983, the number of 'established' engineering consultancies had grown to ten ranging in size from 40 to 420 employees (the Consulting Engineers Council of the Philippines was established in 1976 and affiliated with the FIDIC in 1981). This growth reflected the increased opportunities to plan, design, and supervise the more complex infrastructure projects which underpinned the government's export-oriented industrialization strategy funded by international lending agencies and bilateral loans. Although some of these opportunities have occurred in ad hoc joint ventures it is now the policy of major multilateral agencies, including the World Bank, to only allow local consultants to design specific projects, such as rural highways. This experience has qualified Filipino consultants to work overseas though they do not have sufficient regular commissions to have a direct impact on the structure of the firm in order to make it reach the initial level of Kobayashi's (1981) international scale.

The *DCCD* [Design, Consultation, Construction (Supervision) and Development] *Corporation*, established by university academics in 1957, undertook its first overseas commission in East Pakistan (now Bangladesh) during 1968 (DCCD, 1983). Although the company has developed an advanced matrix organization with ad hoc project teams, its personnel have been engaged almost exclusively to work within the Philippines (for example, projects and services, industrial projects, power generation, and transport projects). Affiliations with foreign firms, however, have broadened its specialities and it hopes to become firmly established on the first rung of Kobayashi's (1981) international scale with work on individual projects in Southeast Asia and the Middle East.

As most Filipino consultants have only fitfully worked overseas it raises the question of how they can attain an international ranking. In a study by Redecon Australia and the DCCD Corporation (Redecon, 1981), however, extensive involvement of Filipino firms is not forecasted, because of the difficulties of providing personnel, at a key staff level, for local consulting services. Many Filipino engineers have already left to work for foreign corporations. Given this situation, how can the nature of government policies be changed to catapult

Filipino consulting firms over the multinational gap? A major fillip to the industry, however, would be the streamlining of government administrative procedures which currently result in delays in payments to consultants of six months or more for completed work. Unlike the contractors, who can delay payments to material suppliers, the consulting firm has to pay for the time of its skilled personnel every two weeks. The practice of delayed payment results in an excessive demand for working capital. This constrains the development of consulting firms as Filipino enterprises do not have the high basic capital nor the access to the public equity market; and borrowing opportunities are also constrained by collateral problems. Until these problems are solved, any movement up Kobayashi's universal scale of internationalization cannot be anticipated. These problems, however, have been less of a barrier for engineering consultancies based in South Korea.

South Korea

An examination of the Korean firms listed in the annual surveys of the *Engineering News-Record* between 1978 and 1985 highlights the fledgling nature of the consulting industry (table 5). No firms made the rankings in 1978 and 1979; only one was ranked in 1980; three were ranked between 1981 and 1984; and four have appeared in the ranking of the 'top 200' consulting firms in 1985. All four firms listed in 1985—Daelim, Daewoo, Hyundai, and Lucky—worked in Malaysia; Daelim, Daewoo and Hyundai also worked in Saudi Arabia and United Arab Emirates; Daelim, Daewoo, and Lucky in Indonesia; Daelim and Daewoo in Iran; and Hyundai and Lucky in Bangladesh. Lucky worked in three countries, Daelim seven, Daewoo nine, and Hyundai ten. Both Hyundai and Lucky had substantial overseas business and appeared in the 'top 100' international firms in 1985. These four internationally-ranked firms were only the vanguard of twenty-five firms registered as international consultants with the Overseas Construction Association of Korea (1983). Of these firms, eighteen were active internationally in that they had job sites outside Korea. Their number included four consulting-contracting firms. This raises the problem of the definition of an international consulting firm in South Korea, particularly as the difficulty was compounded by construction-contracting firms having their own consulting arm. The construction firms involved include Daelim Industrial Co, The Daewoo Corporation, Dong-Ah Construction Ind Co Ltd, Hyundai Engineering and Construction Co Ltd, ICC (Kukje) Construction Co Ltd, Lucky Development Co Ltd, and the Samwhan Corporation.

Daelim Engineering Co Ltd is selected for more detailed analysis because it illustrates this ownership tie-up. The firm, established in 1973, is difficult to locate on Kobayashi's (1981) scale. An early stage of internationalization would

Table 5. Value of foreign billings and overseas percentage of Korean international engineering consultants, 1980-1985 (source: *Engineering News-Record* 13 July 1979; 31 July 1980; 30 July 1981; 29 July 1982; 28 July 1983; 2 August 1984; 1 August 1985; 7 August 1986).

Firm	Value in US$ million (percentage)[a]					
	1980	1981	1982	1983	1984	1985
Chon	–	–	E (92)	D (93)	–	–
Daelim	C (57)	C (69)	C (89)	D (53)	F (23)	E (24)
Daewoo	–	–	–	–	–	E (32)
Hyundai	–	–	F (14)	D (55)	C (59)	D (57)
Lucky	–	–	–	–	C (61)	D (47)

[a] 1980-81: A over $40m; B $20-40m; C $10-20m; D $5-10m; E under $5m.
1982-85: A over $50m; B $30-50m; C $20-29m; D $10-19m; E $5-9m; F under $5m.

have been anticipated as all but 10 of the 103 projects undertaken between 1973 and 1983 were in South Korea (Daelim Engineering Co Ltd, 1983). There was, however, a distinct contrast between work in South Korea and overseas. Within Korea the firm engaged in both civil engineering and plant engineering projects but overseas its activities have been concentrated on the second. This international work, as shown in figure 5, involved the valuable contracts for the design of oil-refining, gas, and petroleum plants for Pertamina in Indonesia and the design of small plants in Saudi Arabia. It also included work on gas projects for Egyptian General Petroleum Corporation under the auspices of the World Bank—design-construct projects that catapulted the firm into prominence among the top international consultants with assets of more than US$13 million, an annual turnover of US$30 million, and a payroll of 600 professional engineers. This emphasis on design-construct probably accounts for the firm's distinctive character. Given the split between overseas and domestic projects one would have expected that the firm would have been at the second level of Kobayashi's scheme. The company, however, has a hybrid structure comprising both a Director of Overseas Marketing and a Vice-President of the Egyptian operations—a feature which points to the difficulties of incorporating, in a meaningful way, organizational structures into Kobayashi's scale.

Daelim's concentration on plant projects when working overseas suggests that the South Korean consulting industry has still to make its name internationally in the more traditional consulting areas such as transport, river catchment and dams, water supply and sewerage, and urban and regional planning. Given the nature and extent of consulting work already being undertaken in South Korea, there is little doubt that the companies are developing these skills in the domestic market. As much stress is placed by international lending agencies on the independence of consultants (that is, that they are not involved as contractors or responsible for manufacturing or product type), moves are already afoot, under the auspices of the Asian Development Bank, to develop the Korean consulting industry in such a way so as to free its ties with the construction industry. This clouded distinction, however, is not marked in Japan where the consulting industry has evolved in a more conventional manner.

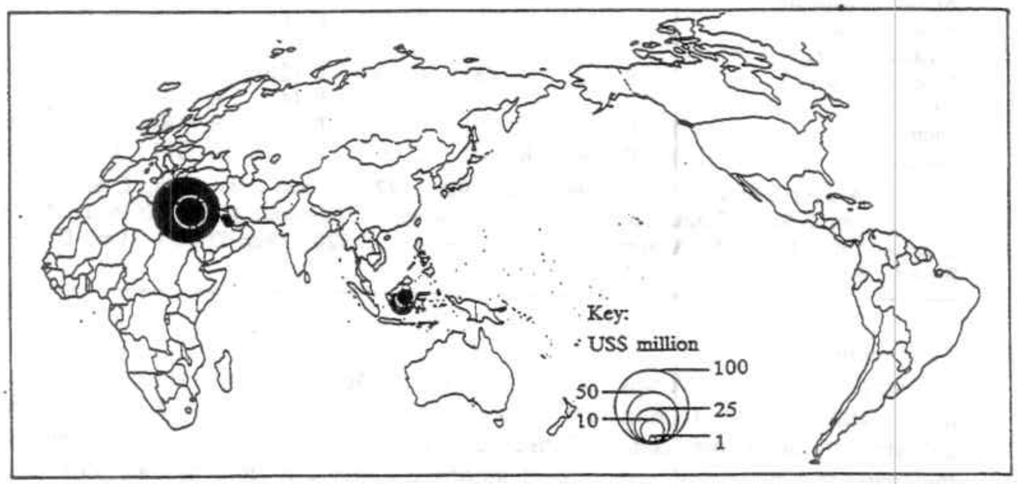


Figure 5. Distribution and value of contracts won by Daelim Engineering Co Ltd, Seoul, South Korea, 1973-1983 (source: Daelim Engineering Co Ltd, 1983).

Japan

Japan's relatively 'weak' international standing in engineering consultancy services, between 1978 and 1985, is reflected in table 6 which shows the value of foreign billings and the percentage of total billings derived from overseas sources. No firm had a complete record for the period under review. Although the number of firms had increased from two in 1978, through three in 1979 and 1980 and eight in 1982, to twelve in 1984 and 1985 these changes, to some extent, reflect the expansion of the rankings to encompass the 'top 200' firms. These firms have been drawn from a small pool of international engineering consultancies (probably less than 100)—a distinct contrast to the large number of domestic firms. This apparent growth in the international standing of consultancy services, however, masks the most striking feature of the rankings of Japanese firms—the intense concentration of overseas business in the hands of two firms. If one is to isolate the reasons for this phenomenon and for Japan's relatively low levels of internationalization one cannot, however, focus attention only on the top firms but one must take a broader view of the industry. This wider perspective involves the consideration of unranked firms and the study of surveys of the industry which have elaborated the international standing of Japanese firms, detailed 'weaknesses' in the industry's structure, and suggested remedies.

Table 6. Value of foreign billings and overseas percentage of Japanese international engineering consultants, 1978–1986 (source: *Engineering News-Record* 13 July 1979; 31 July 1980; 30 July 1981; 29 July 1982; 28 July 1983; 2 August 1984; 1 August 1985; 7 August 1986).

Firm	Value in USS million (percentage)[a]							
	1978	1979	1980	1981	1982	1983	1984	1985
Chuo Kaihatsu	–	–	–	–	F (14)	E (20)	E (23)	E (24)
Electric Power	–	–	–	–	E (100)	E (100)	–	–
Japan Port	–	–	–	–	–	–	F (24)	–
Japan Transport	–	D (30)	–	–	–	–	–	–
Kenzo Tange	–	–	–	–	–	–	F (74)	C (88)
Kisho Kurawa	–	–	–	–	–	–	–	D (72)
Kume	–	–	–	–	–	–	–	E (12)
New Japan	–	–	–	–	–	–	C (49)	C (42)
Nihon Architects	–	–	–	–	–	–	D (29)	E (17)
Nihon Suido	–	–	D (8)	E (10)	E (16)	F (12)	F (9)	E (14)
Nikken Sekkei	–	–	–	–	–	–	D (11)	D (11)
Nippon Koei	–	B (50)	A (48)	A (44)	B (50)	B (49)	B (51)	A (44)
Onoda	–	–	–	–	F (47)	D (77)	C (92)	C (91)
Pacific	B (62)	C (41)	B (36)	B (44)	B (42)	B (49)	C (40)	B (46)
Yachiyo	C (32)	C (49)	B (35)	C (37)	D (23)	B (30)	D (35)	D (30)

[a] 1978: A over S30m; B S10–29m; C S2–9m. 1979: A and B as 1978; C S5–10m; D S2.5–5m; E S1.5–2.5m. 1980–81: A over S40m; B S20–40m; C S10–20m; D S5–10m; E under S5m. 1982–85 A over S50m; B S30–50m; C S20–29m; D S10–19m; E S5–9m; F under S5m.

Case studies

As Nippon Koei and Pacific Consultants are so dominant they are considered in more detail in order to illustrate different facets of Japan's international consulting industry. A third firm, Sanyu Consultants, is also included as it is representative of the firms that have not been ranked among the 'top 200' international consultants. Attention in discussing all three firms is focused on their organizational structure, personnel, and managerial functions. As it is not possible to canvass all facets of these firms, interest is centred on personnel at Sanyu, organizational control at

Pacific International Consultants, and particular management functions at Nippon Koei. If these matters can be dispensed with at this stage, a discussion of the industry as a whole can concentrate on other root causes of Japan's poor international performance and on the intense concentration of business on a few firms.

Sanyu Consultants International Inc (later Sanyu Consultants Inc) was established in Nagoya in 1962. Most of its staff were originally drawn from the Ministry of Agriculture and Forestry—engineers with experience in dams and other hydraulic studies. The company's overseas experience commenced in 1964 when it participated in the Sambor Irrigation Project in Cambodia (part of the Mekong River Basin Development Project) sponsored by the Japanese Government. Its first contract, however, was with the Government of Iran in 1965 for the Taleghan Multipurpose Water Resources Development Project. Since then the company has undertaken many projects throughout Southeast Asia, South Asia, the Middle East, and Africa for the World Bank, Asian Development Bank, and overseas governments, as well as for the Japan International Cooperation Agency and the Overseas Economic cooperation Fund. An analysis of the firm's organizational structure, however, showed that it is still at an early stage of internationalization on Kobayashi's (1981) scale. It maintains its head office in Nagoya and has four regional offices within Japan. Nevertheless, the company has outgrown managing its overseas activities from head office. It has established its overseas projects department in Tokyo. As it is project-based, the company only maintains three regional offices overseas—Bangkok, Cairo, and Seoul—though there are numerous project offices. The projects, as shown in figure 6 have been the basis of training and supervising the company's 252 permanent employees who are sent abroad. The absence of foreign personnel on its staff, however, has constrained the extent to which the company has internationalized.

Pacific Consultants had a rating on Kobayashi's scale which was higher than the rating of Sanyu. This was not surprising as it was established in 1951 as a joint venture with two leading US firms of consultants and local engineers—the first independent consultancy firm to be legally established in Japan. Originally, Pacific Consultants engaged in American military projects, later extending from this base into a joint water-supply venture in the Aichi prefecture in central Japan.

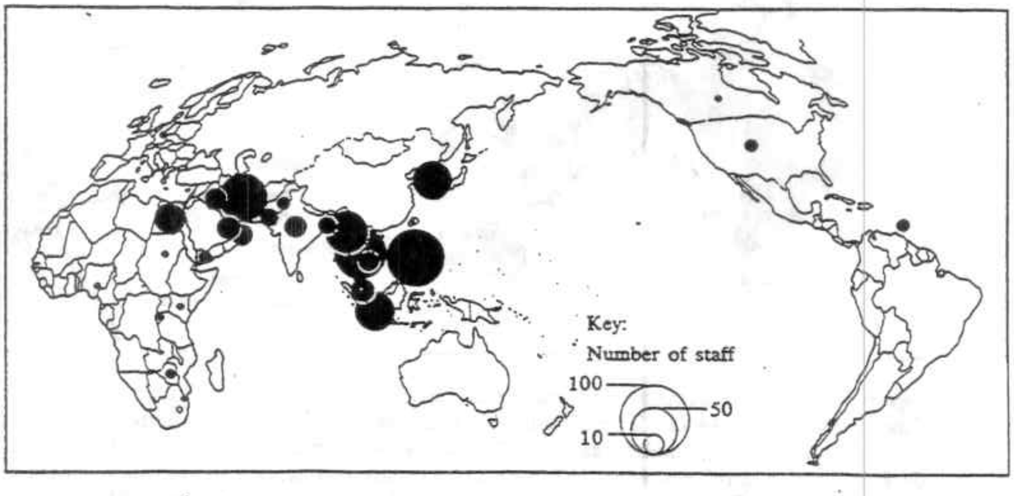


Figure 6. Areas in which the staff of Sanyu Consultants Inc have gained experience, 1964–1983 (Sanyu Consultants Inc, 1983).

Subsequently, the joint venture was amicably dissolved and the Tokyo branch became Pacific Consultants KK, a corporation with 826 employees and still owned by its Japanese cofounders. Since then its geographical spread has exemplified how Japanese consultants have used their domestic experience as a springboard for subsequent overseas projects. As is evident from the list of projects in figure 7, Pacific Consultants have drawn heavily upon their work in designing: national expressways and toll roads in Japan, civil works for the Shinkansen, the Tokyo subway system, Nagasaki airport, flood controls in the Tokyo area, water supply and irrigation, oil and gas facilities, and environmental protection. The nexus between domestic and overseas work should not be overstressed as contracts in Bali have occurred without domestic parallels. These foreign projects have been conducted through a subsidiary, Pacific Consultants International (PCI). Nevertheless, an examination of the company's organizational structure shows that planning of overseas business activities is still conducted within a single department. A full-time director, however, has been appointed to control the overseas business department which operates through ten regional offices that are based largely in Southeast Asia and the Middle East; his authority covers personnel and operational funds.

Nippon Koei, the first engineering firm with an important consultancy base to be established in Japan, was ahead of Pacific Consultants on Kobayashi's internationalization scale as evidenced by its fourteen overseas offices—two more than its rival. Much of its early development, however, was associated with the country's internal postwar recovery. Its founder Kubota Yutaka, and many of its original staff, had, prior to the Pacific War, been engaged in dam building and in electrical power generation in Korea and Manchuria (Manchukuo) (Rix, 1980). They were, therefore, in a position to assist in Tennessee Valley-style projects that were spawned during the American Occupation. As shown in figure 8 the company was also able to take advantage of reparation-funded projects within East and Southeast Asia and able to establish a 'complete' consultancy service in 1954.

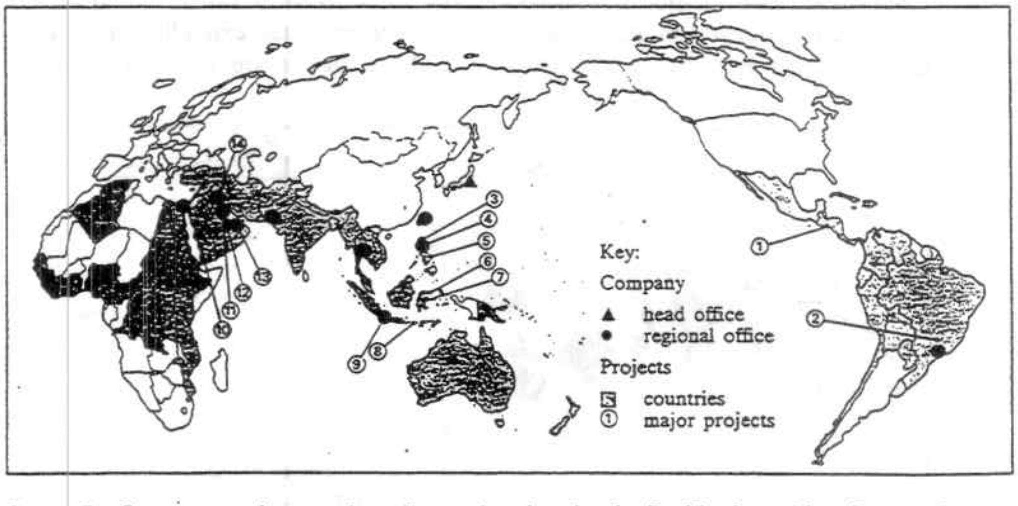


Figure 7. Overseas projects and studies undertaken by the Pacific Consulting Group of Companies, 1951-1982. These include: (1) El Salvador new international airport; (2) Belo Horizonte to Rio de Janeiro railway; (3) Manila rapid transit system; (4) Manila and Laguna de Bay flood control projects; (5) Imelda R Marcos Bridge; (6) Rural roads, North Sulawesi; (7) Aer Tembaga fishing port; (8) the 'Nusa Dua', Bali Tourist Development Plan; (9) new ferry between Java and Sumatra; (10) new road system in Southern and Eastern Regions, Saudi Arabia; (11) Kuwait Hadlan housing project; (12) Kuwait new international airport; (13) Um Qasr container terminal; (14) Baghdad international airport (source: Pacific Consultants, 1983).

Large-scale projects were undertaken in Burma, South Vietnam (Da Nhim Dam as part of the Mekong Project), South Korea, and Indonesia (3-K Dams and Brantas River development). Subsequently, the company has realised its ambition to develop the Asahan project in Indonesia. In the 1960s it extended its activities into Laos and Nepal on aid projects sponsored by the Japanese government and multilateral organizations, such as the United Nations Development Program. During the 1970s further diversification led to the development of projects in Africa and Latin America. The degree of internationalization on Kobayashi's scale, in terms of particular management functions of Japan's largest firm, is variable but, on average, is still at the lower end of the scale. Although the firm has 1300 permanent staff its relations with overseas offices are still bilateral. Decisions, however, are no longer left to a manager in an overseas office but strategies are formulated with the home office. As its engineering research laboratory is based in Japan, head office still has considerable control over research and development. The financing of overseas projects is separated from domestic projects but there is still considerable reliance on concessional finance from the Overseas Cooperation Fund in order to bolster its market penetration.

These brief snapshots of Sanyu Consultants, Pacific Consultants, and Nippon Koei highlight that the management of Japanese engineering consultancies is still at an early stage on Kobayashi's scale compared with the largest American and European firms. Nevertheless, there is no doubting the technical competence of Japanese consultants. Indeed, their skills raise the question of why the Japanese international consultants have been 'weak' competitors.

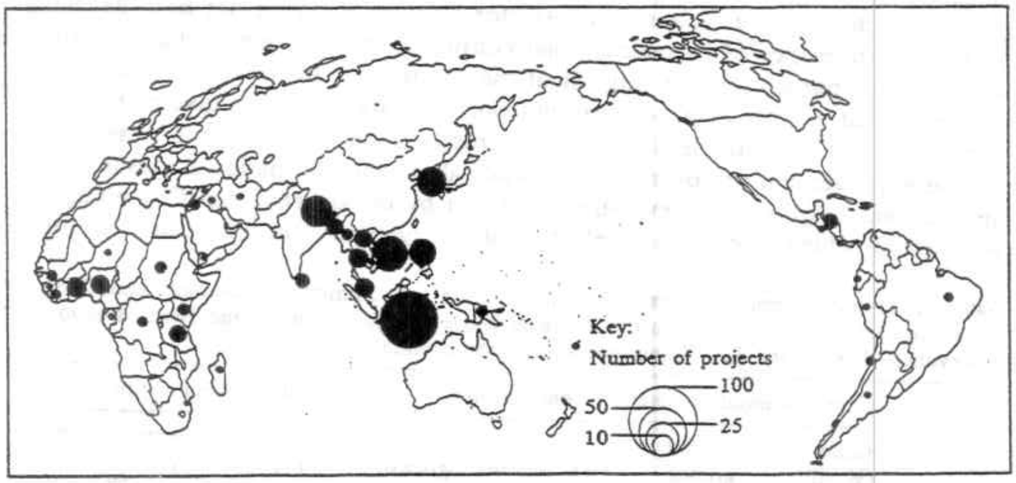


Figure 8. Number of overseas projects undertaken by Nippon Koei Consulting Engineers, 1946-1983 (source: Nippon Koei, 1983).

International performance

As Kobayashi (1981) has highlighted, there is no direct correlation between internationalization and management efficiency. A low level of internationalization does not imply low sales and profits (conversely a high level does not necessarily mean expanded sales or profits). Low levels of internationalization of engineering consultancies and the overall poor performance of these services are, therefore, merely coincidental.

A survey in 1983 of twenty-nine Japanese engineering consultancies by the International Consulting Engineers Association showed that only two firms had made profits in excess of 20% between 1977 and 1982, eight had made less than

10%, and five had made losses. The modal net income before tax was in the range between 1% and 3% in 1980 and 6% in 1981 (Kokusai kensetsu gijutsu kyokai and Nomura sogo kenkyusho, 1983). Japanese government officials, notably from the Ministry of International Trade and Industry, suggested that Japanese services were 'weak competitors' though their performance was not dissimilar to that of leading European firms (for example, Holland's DHV Consulting Engineers and Sweden's SWECO AB and VBB AB).

At first sight, this poor performance is difficult to appreciate because the industry's 13% growth rate has, on occasions, outstripped both gross national product (GNP) and construction output, especially between 1970 and 1980 (table 7). As GNP and construction output were tapering off in the early 1970s from high levels of growth, the performance of the consulting industry has been discounted by government officials. Indeed, confidence in the industry has been lacking as surges in growth have been offset by marked reverses. Whereas the construction industry consistently provided over 20% of GNP between 1972 and 1982, the consulting industry's performance has fluctuated between 0.009 and 0.023%—a poor showing compared with 0.04% for West Germany, 0.08% for France, and 0.14% for the United Kingdom in 1981.

More disturbing to Japanese government officials has been that the contribution of international business to the consulting industry's total revenue between 1972 and 1981 has only fitfully matched the growth of the domestic industry (table 8). Admittedly, the number and value of projects handled by the domestic industry have been buoyed by contributions from the public sector, notably from the six largest cities, the provincial governments, the public corporations, and the Ministry of Construction. Nevertheless, public-sector business in Japan has been declining and the performance of its international consulting industry against that yardstick is discouraging, particularly for officials at the Ministry of International Trade who recognise that exports of goods may not only give way to direct investment and production but may also lead to an even greater boom in the export of services.

The apprehension of government officials about the international consultancy services was heightened by evidence produced by the International Consulting Association which showed that only two out of nineteen correspondents reported

Table 7. Trends in gross national product, domestic construction and overseas consulting by Japanese firms, 1972–1982 (source: Kokusai kensetsu gijutsu kyokai and Nomura sogo kenkyusho, 1983, page 39; Kaigai kensetsu kihon mondai kentokai, 1982).

Year	Gross national product		Construction		Overseas consulting		
	value (¥ billion)	annual growth rate (%)	value (¥ billion)	annual growth rate (%)	value (¥ billion)	annual growth rate (%)	ratio of GNP (%)
1972	96884	16.5	21463	28.7	8.6	−13.8	0.009
1973	117258	21.0	28667	33.6	10.3	20.1	0.009
1974	138045	18.4	29394	2.5	14.2	37.1	0.010
1975	151797	10.0	31624	7.6	17.7	24.6	0.012
1976	170290	12.2	34197	8.1	25.5	44.2	0.015
1977	188804	10.9	38799	13.5	42.8	63.0	0.023
1978	206763	9.5	42686	10.0	26.3	−38.0	0.013
1979	221966	7.4	47922	12.3	22.7	−13.8	0.010
1980	238894	7.7	49700	3.7	25.7	13.4	0.011
1981	251536	5.2	50700	2.0	42.9	71.9	0.018
1982	266000	5.8	53770	6.1	na[a]	na	na

[a] na means not available.

more than 70% of their income from overseas sources (Kokusai kensetsu gijitsu kyokai and Nomura sogo kenkyusho, 1983). The sample population was also heavily dependent for business on the Overseas Economic Cooperation Fund (42%) and the Japanese International Cooperation Agency (16%). Conversely, only a small contribution was made by the World Bank (5%), by the Asian Development Bank (3%) and by voluntary organizations (3%). As private organizations made up the balance it was disturbing to government officials that the private contribution had declined from its 1976 peak to leave the services heavily dependent on the

Table 8. Number and average value of domestic and international projects undertaken by Japanese consultants, 1970–1981 (source: Kokusai kensetsu gijitsu kyokai and Nomura sogo kenkyusho, 1983, page 41).

Year	Domestic projects			International projects			
	number	total value (¥ million)	average value (¥ million)	number	total value (¥ million)	average value (¥ million)	share of total value (%)
1970	16531	24837	1.502	126	7710	61.190	23.7
1971	17396	31050	1.785	176	9982	56.715	24.3
1972	23791	47971	2.016	176	8600	48.864	15.2
1973	26406	62473	2.366	140	10330	73.786	14.2
1974	31438	84101	2.675	143	14162	99.035	14.4
1975	29899	96921	3.242	182	17651	96.984	15.4
1976	34417	103818	3.016	197	25458	129.228	19.7
1977	35382	118379	3.346	293	42780	146.007	26.5
1978	40161	152474	3.797	231	26287	113.796	14.7
1979	43057	185453	4.307	507	22651	44.677	10.9
1980	48045	194224	4.043	453	25683	56.695	11.7
1981	50639	202300	3.995	332	44155	132.996	17.9
1982	na[a]	na	na	350	41150	117.571	na

[a] na means not available.

Table 9. Sources of earnings for consultancy services provided by Japanese consultants, 1972–1985 (source: Kokusai kensetsu gijitsu kyokai and Nomura sogo kenkyusho, 1983; Keikakukyoku kokusaika, 1983a; 1983b; Kensetsusho, 1985; 1986).

Year	Public sector		Private sector[a]		Total ¥ million
	¥ million	percentage of total	¥ million	percentage of total	
1972	2408	28.0	6192	72.0	8600
1973	2063	20.0	8264	80.0	10327
1974	2702	19.1	11460	80.9	14162
1975	4303	24.4	13348	75.6	17651
1976	9240	36.3	16218	63.7	25458
1977	19484	45.5	23296	54.5	42780
1978	14719	56.0	11568	44.0	26287
1979	10826	47.8	11825	52.2	22651
1980	13924	54.2	11759	45.8	25683
1981	35387	80.1	8768	19.9	44155
1982	32199	78.2	8951	21.7	41150
1983	38353	74.0	13498	26.0	51851
1984	35068	87.2	5131	12.8	40217
1985	30354	84.3	5650	15.7	36004

[a] Private sector includes overseas affiliates.

Overseas Economic Cooperation Fund. The dilemma is demonstrated even more starkly in table 9, in which is shown a heavy reliance on the 'public sector' (that is, Overseas Economic Cooperation Agency, Japan International Cooperation Agency, and non-profit-making organizations). More disturbing for government officials has been that this dependence has increased markedly since 1972, and accounts for almost 88% of the value of all projects in 1984.

This 'weakness' was aggravated for government officials by the intense structural concentration within the industry. The 'top ten' firms, for example, were responsible for 86% of all overseas business. They also accounted for 67% of business from Japan International Cooperation Agency, 92% from the Overseas Economic Cooperation Fund, 96% from private firms, 98% from the Asian Development Bank, and 100% from the World Bank. This feature was common throughout all branches of international consulting—the 'top five' accounting for 83% of planning and investigation studies, 84% of design work, and 94% of construction management.

The problem of structural concentration was compounded for Japanese government officials by spatial concentration. As is shown in table 10, the export of services was heavily dependent on the Asian region. The Middle East has faded as a market from its high point in 1977. Although Africa showed promise in 1979 and 1983, as did Latin America in 1980 and 1985, these upsurges were not sustained to mitigate the increasing reliance on Asian markets.

As highlighted by an examination of the 'top ten' markets for Japanese consultants, between 1978 and 1985, there was a heavy dependence on four countries which belong to the Association of Southeast Asian Nations (ASEAN)—Indonesia, Malaysia, the Philippines, and Thailand (table 11). Indonesia was the undisputed market leader throughout the period under review and accounted for between one-fifth and two-fifths of all business, with the Philippines, Malaysia, and Thailand attracting consistent business. In the Middle East, Saudi Arabia was prominent until 1980 and Iraq until the onset of the Gulf War. These more or less stable features of the rankings have to be contrasted with unstable characteristics. Between 1978 and 1985 South Korea, Libya, Burma, China, Kenya, and Bangladesh enjoyed fitful prominence and Egypt, Nigeria, Sri Lanka, and Pakistan appeared in the rankings twice. Twelve countries, including Nepal, Gabon, New Zealand, Dominica, Honduras, and Syria, appeared only once.

Part of this instability has been attributed by government officials to the nature of the industry during the late 1970s and early 1980s. As shown in table 12, between one-third and one-half of the overseas fees earned by the industry have

Table 10. Percentage regional distribution of international contracts won by Japanese consultants, 1977–1985 (source: Kensetsusho kensetsu keizai kyoku kokusaika, 1985).

Year	Region				
	Asia	Middle East	Africa	Latin America	other
1977	48	40	4	4	4
1978	60	29	7	3	1
1979	56	19	15	9	1
1980	58	24	7	10	2
1981	71	12	7	8	2
1982	72	17	8	3	0
1983	66	15	11	5	3
1984	80	10	4	5	1
1985	73	7	6	13	1

been derived from power and transport projects; fees from telecommunication and social welfare have also been steady but their contribution has been markedly smaller. In contrast, fees from commercial, mining, and industry projects have fluctuated wildly. Returns from irrigation and agriculture have shown fitful promise but planning has not been an important field overseas.

Table 11. 'Top ten' overseas markets for Japanese consultants, 1978–1985 (source: Keikakukyoku kokusaika, 1983a; 1983b; Kensetsu keizai kyoku kokusaika, 1985; 1986).

Country	1978	1979	1980	1981	1982	1983	1984	1985
Indonesia	1	1	1	1	1	1	1	1
Iraq	2	6	2	3	3	–	–	–
Saudi Arabia	3	2	–	–	–	–	–	–
Korea	4	–	5	–	–	–	4	3
Libya	5	5	–	7	–	–	–	–
Philippines	6	3	3	2	2	4	3	4
Malaysia	7	7	6	6	6	3	8	10
Thailand	8	4	4	4	4	5	2	5
Iran	9	–	–	–	–	–	–	–
China	10	–	–	–	–	–	10	2
Bolivia	–	8	–	–	–	–	–	–
Burma	–	9	–	8	–	6	–	7
Kenya	–	10	–	9	8	–	–	–
Egypt	–	–	7	–	–	–	5	–
Gabon	–	–	8	–	–	–	–	–
Bangladesh	–	–	9	–	–	10	7	–
Nigeria	–	–	10	–	–	9	–	–
Paraguay	–	–	–	5	–	–	–	–
Sri Lanka	–	–	–	10	–	–	–	8
Nepal	–	–	–	–	5	–	–	–
Pakistan	–	–	–	–	7	–	6	–
Tanzania	–	–	–	–	9	–	–	–
Jordan	–	–	–	–	10	–	–	–
Kuwait	–	–	–	–	–	2	–	–
New Zealand	–	–	–	–	–	7	–	–
Dominica	–	–	–	–	–	8	–	–
Singapore	–	–	–	–	–	–	9	–
Honduras	–	–	–	–	–	–	–	6
Syria	–	–	–	–	–	–	–	9

Table 12. Percentage fees earned by field of consultancy for services provided by Japanese firms, 1977–1982 (source: Kokusai kensetsu gijutsu kyokai and Nomura sogo kenkyusho, 1983).

Field	1977	1978	1979	1980	1981	1982
Power	15.0	34.6	16.0	23.7	31.3	22.2
Transport	22.3	21.0	23.8	22.8	36.6	15.6
Telecommunications	7.2	5.2	8.2	7.6	3.5	7.3
Irrigation, etc	2.6	4.8	9.9	0.2	4.1	9.7
Agriculture	0.0	4.4	10.8	3.6	0.7	2.7
Mining and industry	21.3	4.6	0.0	18.1	10.4	12.4
Social welfare	2.2	4.4	5.6	9.8	4.5	2.7
Planning	3.0	0.0	0.0	1.1	1.6	0.8
Commerce	26.5	21.1	25.8	13.0	7.6	26.1
Other	0.0	0.0	0.0	0.0	0.0	0.0
Total	100.0	100.0	100.0	100.0	100.0	100.0

Note: Due to rounding figures may not total 100.0.

Participants engaged in trading services overseas have attributed their competitive weakness to the government's lack of encouragement of bilateral aid, to limited human resources, and to inability to cope with internationalism (*kokusaika*). Mitigating circumstances have been the short life-span of engineering consultancy, as it was essentially a post-Pacific War phenomenon. The problems of its comparative youthfulness, however, have been compounded by the limited use of consultants by Japanese enterprises. The use of in-house consultants by the government organizations, for instance, deprived them of the experience necessary for working overseas. Weakness in the government's strategy of financing projects compounded these problems. It is, however, considered to be of secondary importance to the lack of ability or willingness to work internationally. These mitigating circumstances, however, do not offset what is perceived to be the root cause of Japanese international consulting falling below expectations—the poor linguistic capacity of Japanese consultants. There is a pressing need to overcome these restraints, as the recession in the domestic market has pushed the general construction (building and civil works) industry into overseas markets and has increased the demand for international consultants to identify upcoming business opportunities. Industry associations, notably the Overseas Construction Association (*Kaigai kensetsu kyokai*), International Consulting Engineers Association (*Kokusai kensetsu gijitsu kyokai*), International Development Centre (*Kokusai kaihatsu senta*), and Engineering Consulting Firms Association (*Kaigai konsarutingu kigyo kyokai*), have been working to establish the industry's credibility and provide a springboard for exporting its services (see, for example, Kaigai kensetsu kyokai et al, 1983a; 1983b).

The Engineering Consulting Firms Association (ECFA), founded in 1964 with fourteen members, has been the main spearhead for consultants. Founded on the presumption that it would be supported by government ministries, ECFA has been the main funnel for the bulk of state assistance (Kaigai konsarutingu kigyo kyokai, 1984). In particular, ECFA has promoted research into the transfer of appropriate technology and on Japanese experience in the development of small-scale and medium-scale industries. It has also despatched volunteer survey teams and others at its own expense (though with a subsidy from the Ministry of International Trade and Industry), and has coordinated large-scale projects, such as the Baghdad Master Plan, involving seven consulting firms.

As project economists have been in short supply, ECFA has also lent out its staff to plug gaps, especially where negotiations for a contract are likely to be hamstrung by the lack of these skills or by an inability to communicate in English. Since 1973, some 2000 missions have been despatched by ECFA to identify projects and to lobby over seventy developing countries, in the hope that one of its members will secure the contract. Between its inception in 1964 and 1982 ECFA has been successful in securing 174 contracts for its members. Individual firms have augmented ECFA's efforts by sending out their own missions, as aid projects are a means of profit and future business opportunities. These activities appeared to have produced results, as the overall contribution of Japanese international consultants had increased from 3% of foreign billings in 1982, through 4.5% in a declining market in 1984, to 6.2% in 1985; Korean consultants also captured more than 1% of these billings in both 1984 and 1985 (*Engineering News-Record* 1 August 1985; 7 August 1986). A further improvement in the position of Japanese consultants can be anticipated, as the Japanese government intends to double its Official Development Assistance (ODA) to Southeast Asia between 1986 and 1993 (*Engineering News-Record* 7 August 1985). No doubt American-based and European-based consultants will be seeking to pool their skills in joint ventures with their Japanese counterparts.

Conclusions

A study of this nature cannot come to any firm conclusions without a more detailed and thorough survey of engineering consultancies, between countries and over time. At best, it points to some of the differences between international engineering consultancies which are headquartered in different countries. These differences are not taken to reflect the variations in 'management culture' that are rooted in the distinctive traits of differing societies (for example, United States, Europe, and Japan)—an ethnocentric approach. Instead, they are seen as representing differing degrees of progress on Kobayashi's (1981) universal scale of internationalization which is based on the expansion of a management perspective from a local to a global scale and on the degree to which a firm's international operations are integrated into its overall organization—a geocentric approach. Differences between the degree of internationalization attained by particular Japanese, Korean, and Filipino engineering consultancies, and the higher levels reached by larger American and European global organizations probably reflect their latecomer status. An analysis of these firms in terms of their organizational control, personnel, and particular management functions highlight that they have only recently moved towards internationalization and are still at a relatively early stage.

The problem faced by Asian engineering consultancies, however, may not be as smooth as Kobayashi's scheme suggests. A move from head office, through delegation of management to overseas units involves a shift from undertaking isolated, individual overseas projects to an incorporation of them as a recurrent activity of an international department. In turn, the shift from regional coordination to management from a global perspective involves a change from considering overseas activities as a regular feature of the short-term planning to an absorption of them into the long-term planning of the total enterprise. Each change represents a financial hurdle, from raising funds on an individual project basis to a single comprehensive system in which funds are derived from both domestic and foreign capital markets. The key to managerial efficiency in this process, however, is not related to the degree of internationalization. According to Kobayashi (1981, page 11) it probably has more to do with the location of head office and the degree to which internationalization has permeated the whole of a company's organization. It is these assertions that a more thorough study will have to address.

Acknowledgements. In writing this paper the following are acknowledged: Professor J A Black and Dr Leone Yandell (Department of Transport Engineering, University of New South Wales); Consulting Engineers Council of the Philippines (Jaime A Cura); DCCD Engineering Corporation (Eric R Ruiz); Daelim Engineering Co Ltd (J H Cho, Dae Il Kim); Overseas Construction Association of Korea (Ki Cheol Kim, Kiong Wha Park, Myong Huh); Engineering Consulting Firms Association (Ozaki Takao, Yamaguchi Hitoaki, Wagatsuma Misako); International Engineering Consultants Association (Tokuhiro Hideo); Nippon Koei (Kinoshita Nobuaki); Pacific Consultants (Kawano Yasuo); Sanyu Consultants (Fukuda Eiji, Higuchi Shichiro, Miyazaki Yasushi); and Barbara Banks. The figures were drawn by Keith Mitchell, Nigel Duffey, and Ian Heyward (Cartographic Section, Department of Human Geography, Research School of Pacific Studies, the Australian National University). Some drafting was also undertaken by Joe Rimmer.

References

Armytage W H G, 1976 *A Social History of Engineering* (Faber and Faber, London)
Association of Consulting Engineers, 1970, "Overseas work entrusted to members during 1969", The Association of Consulting Engineers, Alliance House, Caxton St, London SW1
Bell S P, 1975 *A Bibliographical Index of British Engineers in the 19th Century* (Garland, New York)

Bryant E K, 1973, "Opportunities for American engineers in overseas work" *Engineering Issues—Journal of Professional Practice, Proceedings of the American Society of Civil Engineers* **99** 433–440

Cardinal R J, 1980, "Matrix management may meet your needs for the 80s" *Issues in Engineering—Journal of Professional Practice, Proceedings of the American Society of Civil Engineers* **106** 27–39

Civil Engineering Economic Development Committee, 1978 *Design and Export: The Better Use of Design Resources for Meeting the Needs of UK Clients and Earning Foreign Currency* National Economic Development Office, 21 Millbank, London SW1, England

Clarke I M, 1984 *The Spatial Organisation of Corporations: A Case Study of the Multinational Chemicals Industry with Specific Reference to ICI* unpublished PhD thesis, Department of Human Geography, Australian National University, PO Box 4, Canberra, ACT 2601, Australia

Consulting Engineer, 1985 *Consulting Engineers Who's Who and Yearbook* (Municipal Publications Ltd, London)

Consulting Engineering Services, 1971 *Export Potential of Indian Civil Engineering Consultancy and Construction and Associated Services and Supplies* Consulting Engineering Services (India) Private Ltd, New Delhi, India

Daelim Engineering Co Ltd, 1983 *DEC Capability* Daelim Engineering Co Ltd, 1-423, Yoido-dong, Yomgdeungpo-ku, CPO Box 7035, Seoul, South Korea

DCCD, 1983 *25 Years of Consultancy* Design, Consultation, Construction (Supervision) and Development Corporation, Sol Building, 5175 Amorsolo St, Legazpi Village, Makati, Metro Manila, The Philippines

Engineering News-Record 1979–1986 *Engineering News-Record* (The McGraw-Hill Construction Weekly) (McGraw-Hill, New York)

Enis B M, Roering K J, 1981, "Services marketing: different products: different strategy", in *Marketing of Services* Eds J Donelly, W George (American Marketing Association, Chicago) pp 1–4

ESCAP, 1977 *Consultancy Services Available in Developing ESCAP Countries* Economic and Social Commission for Asia and the Pacific, United Nations Building, Rajdamnern Avenue, Bangkok 2, Thailand

Fucik E M, 1964, "Consulting engineering practice overseas" *Journal of Professional Practice—Proceedings of the American Society of Civil Engineers* **90** 37–43

Guttman H P, 1976 *The International Consultant* (McGraw-Hill, New York)

Kaigai kensetsu kihon mondai kentokai, 1982, "Kaigai kensetsu shinko no kihon tekihokoku (chukan tori matome) [Basic directions and methods established for the provision of overseas construction (interim report)]", Kaigai kensetsu kihon mondai kentokai, International Affairs Division, Ministry of Construction, 2-1-3, Kasumigaseki, Chiyoda-ku, Tokyo, Japan 100

Kaigai kensetsu kyokai, Nikko Risa-chi Senta, 1983a, "Kaigai kensetsu no kihon tekihoko to taiosaku (II): risuku taisaku no sogoteki kenkyu (Countermeasures and basic directions for overseas engineering: study of comprehensive countermeasures for risk)", Kaigai kensetsu kyokai, 8th floor, Tokyo Tatemono Daisan, Yaesu Building, 1-9-1, Hatchobori, Chuo-ku, Tokyo, Japan 104

Kaigai kensetsu kyokai, Nomura sogo kenkyusho, 1983b, "Kaigai kensetsu no kihon tekihoko to taioaku (III): Kaigai kensetsu no yoseika ni kansuru chosa kenkyu (Countermeasures and basic directions for overseas engineering study and survey concerning training measures for overseas engineering)", Kaigai kensetsu kyokai, 8th floor, Tokyo Tatemono Daisan, Yaesu Building, 1-9-1, Hatchobori, Chuo-ku, Tokyo, Japan 104

Kaigai konsarutingu kigyo kyokai, 1982, "Kaihatsu tojo kuni no konsarutingu kigyo ni kansuru chosa: genjo to tenbo (Survey of the consulting industry in developing countries: observations and existing state of affairs)", Kaigai konsarutingu kigyo kyokai, Room 3210, PO Box 60, Kasumigaseki Building, 2–5 Kasumigaseki, 3-chome, Chiyoda-ku, Tokyo, Japan 100

Kaigai konsarutingu kigyo kyokai, 1984, "Aru dantai no kiroku: kaigai konsarutingu kigyo kyokai ninenshi (One organization's record: a twenty year history of the Engineering Consulting Firms Association)", Kaigai konsarutingu kigyo kyokai, Room 3210, PO Box 60, Kasumigaseki Building, 2–5 Kasumigaseki, 3-chome, Chiyoda-ku, Tokyo, Japan 100

Keikakukyoku kokusaika, 1983a, "Kaigai kensetsu koji oyobi kensetsu kanren gyomu juchu jisseki (Past results of acceptance of orders as well as overseas construction work)", Keikakukyoku kokusaika, International Affairs Division, Ministry of Construction, 2-1-3 Kasumigaseki, Chiyoda-ku, Tokyo, Japan 100

Keikakukyoku kokusaika, 1983b, "Kokusai kyoryoku no genkyo (Present state of international cooperation", Keikakukyoku kokusaika, International Affairs Division, Ministry of Construction, 2-1-3 Kasumigaseki, Chiyoda-ku, Tokyo, Japan 100

Kensetsu keizai kyoku kokusaika, 1985, "Kokusai kyoryoku no genkyo (Present state of international cooperation)", Kensetsu keizai kyoku kokusaika, International Affairs Division, Ministry of Construction, 2-1-3 Kasumigaseki, Chiyoda-ku, Tokyo, Japan 100

Kensetsu keizai kyoku kokusaika, 1986, "Kokusai kyoryoku no genkyo (Present state of international cooperation), Kensetsu keizai kyoku kokusaika, International Affairs Division, Ministry of Construction, 2-1-3 Kasumigaseki, Chiyoda-ku, Tokyo, Japan 100

Kensetsusho, 1985, "Showa 59 nendo kaigai kensetsu kojito juchu jisseki (Past results of contracts accepted by the overseas construction industry etc, 1984)", Kensetsusho, International Affairs Division, Ministry of Construction, 2-1-3 Kasumigaseki, Chiyoda-ku, Tokyo, Japan 100

Kensetsusho, 1986, "Showa 60 nendo kaigai kensetsu kojito juchu jisseki (Past results of contracts accepted by the overseas construction industry etc, 1985", Kensetsusho, International Affairs Division, Ministry of Construction, 2-1-3 Kasumigaseki, Chiyoda-ku, Tokyo, Japan 100

Kensetsusho kensetsu keizai kyoku kokusaika, 1985, "Kaigai kensetsu kihon mondai kentokai (Basic direction and methods established for the promotion of overseas construction)", Kensetsusho kensetsu keizaiku kokusaika, International Affairs Division, Ministry of Construction, 2-1-3 Kasumigaseki, Chiyoda-ku, Tokyo, Japan 100

Kobayashi N, 1981, "The present and future of Japanese multinational enterprises: a comparative analysis of the Japanese and US-European multinational management" *The Wheel Extended* **9** 10-18

Kokusai kensetsu gijutsu kyokai, Nomura sogo kenkyusho, 1983, "Kensetsu konsarutanto no kaigai katsudo sokushin no tameno kihon mondai to choki bijyon sakutei no kenkyu (A study for consulting engineers of the promotion of their overseas activities and related long-term problems)" Kokusai kensetsu kyokai, International Engineering Consultants (Japan), 5-3-23 Kojimachi, Chiyoda-ku, Tokyo, Japan 102

Kollar K L, 1963, "Exporting consulting engineering" *Civil Engineering* July issue, pp 58-59

Kuecken J A, 1978 *Starting and Managing your Own Engineering Practice* (Van Nostrand, New York)

Levy R M, 1980 *The Professionalization of American Architects and Civil Engineers, 1865-1917* unpublished PhD thesis in Architecture in the Graduate Division of the University of California, Berkeley, CA 94720, USA

Lindstrom V G Jr, 1982, "CM's expanding market" *Constructor* September, pp 24-27

Locke R R, 1984 *The End of the Practical Man: Entrepreneurship and Higher Education in Germany, France and Great Britain, 1880-1940* (Jai Press Inc, Greenwich, CT)

Logcher R D, Levitt R E, 1979, "Organization and control of engineering design firms" *Issues in Engineering—Journal of Professional Activities, Proceedings of the American Society of Civil Engineers* **105** 7-13

Morgan P R, Kamil M H, 1982, "Contract arrangements in the Third World" *Issues in Engineering—Journal of Professional Activities, Proceedings of the American Society of Civil Engineers* **108** 243-250

Narver D L, 1975, "The adopted ASCE guide to turnkey construction contracts" *Issues in Engineering—Journal of Professional Activities, Proceedings of the American Society of Civil Engineers* **101** 235-241

Nippon Koei, 1983 *Nippon Koei Consulting Engineers* Nippon Koei, 4 Kojimachi, 5-chome, Chiyoda-ku, Tokyo, Japan

Noyelle T J, Stanback T M Jr, 1984 *The Economic Transformation of American Cities* (Rowman and Allanheld, Totowa, NJ)

OEEC, 1958, "Some aspects of consultant engineering in the United States", Project 357, Organisation for European Economic Cooperation, European Productivity Agency, 2 rue André-Pascal, Paris 16, France

Overseas Construction Association of Korea, 1983 *List of Member Companies 1983* Overseas Construction Association of Korea, Room 1612, Kuk-dong Building, 3ka Chungmu-ro, CPO Box 8543, Seoul, South Korea

Pacific Consultants, 1983 *The Pacific Consultants Group of Companies* Pacific Consultants, 8-2, 2-chome, Jingumae, Shibuya-ku, Tokyo 150, Japan

Pinnell S S, 1980, "Construction/engineering management: a comparison" *Issues in Engineering—Journal of Professional Activities, Proceedings of the American Society of Civil Engineers* **106** 405–413

Redecon, 1981, "Final report on the National Construction Industry", Redecon Australia and DCCD Engineering Corporation, Ministry of Public Works and Highways, Bonifacio Drive, Port Area, Metro Manila, The Philippines

Registry of Scientific and Technical Services, 1973, "Engineering Consulting firms in the Asian and Pacific Region", Registry of Scientific and Technical Services for the Asian and Pacific Region, PO Box 62, Jamison Centre, Canberra, ACT 2614, Australia

Riddle D I, Sours M H, 1984, "Service industries as growth leaders on the Pacific Rim" *Asia Pacific Journal of Management* **1**(3) 190–199

Rimmer P J, 1980, "The British expert cometh: the imprint of the individual on Australia's economic landscape", in *Of Time and Place* Eds J N Jennings, G J R Linge (Australian National University Press, Canberra) pp 48–71

Rimmer P J, 1984, "Consultancy services: supply to Southeast Asia from Australia", ASEAN–Australia Economic Papers 7, ASEAN–Australia Joint Joint Research Project; c/o Economic Planning Unit, Prime Minister's Department, Kuala Lumpur, Malaysia; or c/o Research School of Pacific Studies, Australian National University, Canberra, ACT 2601, Australia

Rimmer P J, 1986, "Australian and Japanese international engineering consultancy services", Pacific Economic Papers 131, Australia–Japan Centre, Research School of Pacific Studies, Australian National University, Canberra, ACT 2601, Australia

Rimmer P J, Black J A, 1983, "Global financiers, consultants and contractors in the Southwest Pacific since 1970" *Pacific Viewpoint* **24** 112–139

Rimmer P J, Black J A, 1984, "Japanese, Korean and Filipino international construction contractors: structure and performance", WP-40, Development Studies Centre, The Australian National University, PO Box 4, Canberra, ACT 2601, Australia

Rix A, 1980 *Japan's Economic Aid: Policy-making and Politics* (Croom Helm, Beckenham, Kent)

Sanyu Consultants Inc, 1983 *Company Qualification and Experience*, Tokyo Sanyu Consultants Inc, 15/22, 2-chome, Nishiki, Naka-ku, Nagoya, Japan

Sassen-Koob S, 1986, "New York City: economic restructuring and immigration" *Development and Change* **17** 85–119

Seow G, Tucker K, Sundberg M, 1984, "ASEAN–Australia trade in tourist services", ASEAN–Australia Economic Papers 11, ASEAN–Australia Joint Joint Research Project; c/o Economic Planning Unit, Prime Minister's Department, Kuala Lumpur, Malaysia; or c/o Research School of Pacific Studies, Australian National University, Canberra, ACT 2601, Australia

Stanback T M Jr, 1979 *Understanding the Service Economy: Employment, Productivity, Location* (The Johns Hopkins University Press, Baltimore, MD)

Stanback T M Jr, Bearse P J, Noyelle T J, Karasek R A, 1981 *Services: The New Economy* (Rowman and Allanheld, Totowa, NJ)

Stanback T M Jr, Noyelle T J, 1982 *Cities in Transition: Changing Job Structures in Atlanta, Denver, Buffalo, Columbus (Ohio), Nashville, Charlotte* (Allanheld, Osman, Totowa, NJ)

Stanley C M, 1961 *The Consulting Engineer* (John Wiley, New York)

Tucker K A, 1981, "Traded services in the world economy", WP-16, Bureau of Industry Economics, Canberra, Australia

Tucker K A, Seow G, Sundberg M, 1983, "Services in ASEAN–Australia Trade", ASEAN–Australia Economic Papers 2, ASEAN–Australia Joint Joint Research Project; c/o Economic Planning Unit, Prime Minister's Department, Kuala Lumpur, Malaysia; or c/o Research School of Pacific Studies, Australian National University, Canberra, ACT 2601, Australia

US Department of Commerce, 1970 *Engineers' Overseas Handbook* Bureau of Domestic Commerce (US Government Printing Office, Washington, DC)

© 1988 a Pion publication printed in Great Britain

70

WHARTON REPROGRAPHICS

Lawyers Abroad:

THE INTERNATIONALIZATION OF LEGAL PRACTICE

LIPPINCOTT LIBRARY
DOCUMENT DELIVERY

Debora L. Spar

Since the late 1960s, a number of U.S. and British law firms have quietly, successfully, gone global. Traipsing after their far-flung corporate clients, these law firms have slowly but steadily established their own global networks. Together, they employ thousands of people and have created management structures that control and coordinate expanding clusters of relationships. Several of the largest law firms are multinationals in their own right—but with one critical distinction. Unlike the manufacturing firms that preceded them abroad, law firms are essentially service firms, selling the ephemeral products of information, skills, and advice. As such, they rank among the most important global providers of information-based services, and they have had to grapple with many of the issues facing other information-based firms. Law firms have not always been successful in their international expansion, and most of them still concentrate on the domestic aspects of their work. Several major international law firms, however—such as White & Case, Cleary, Gottlieb, and Clifford Chance—have built organizations whose strengths extend considerably beyond their home bases.

I am tremendously grateful for the research assistance of Elizabeth Stein and the financial assistance provided by the Division of Research, Harvard Business School. Earlier drafts benefited greatly from the comments of Sven Beckert, I. J. Alexander Dyck, Willis Emmons, Joseph Hinsey, George Lodge, Tony Oettinger, Lynn Paine, Anne-Marie Slaughter, Susan Strange, Richard Tedlow, Richard Vietor, Lou Wells, and David Yoffie. Many attorneys, who wish to remain anonymous, also contributed to earlier versions and shared with me their insights about the practice of law

Background: The Evolution of Legal Practice in the United States

Like accountants, doctors, consultants, and advertising agencies, lawyers have always sold a somewhat awkward product. They sell an information-based service—a product whose value lies in its customization, a product that is difficult to stockpile or resell, and a product based inherently on human, rather than physical, capital. Unlike many service businesses, the services they sell vary substantially from one customer to the next. Essentially, law firms sell the promise of a product—the promise that they will create a contract, or a will, or a business arrangement that best suits their customer's needs.[1]

For a long time, these distinctive characteristics set law apart from the rugged world of commerce. In the early years of the United States, lawyers were a small and highly educated minority,[2] blessed with a distinct social status that Alexis de Tocqueville attributed to "the special information which lawyers derive from their studies."[3] The possession of this information allowed lawyers to perform a wide range of services; they argued cases, advised clients, copied documents, and delivered papers personally. During this heyday of independent practice, advocacy before the bench was viewed as the lawyer's core competency, and most lawyers worked as independent and all-purpose practitioners.

Between 1870 and 1890, however, a flood of commercial legislation transformed the legal landscape, and technological innovations (such as the typewriter and the telegraph) revolutionized the tools of the trade.[4] Consequently, legal practice experienced some of the consolidation mania that was then sweeping across American industry. Two-person partnerships rapidly gave way to four, five, and six-member firms. And lawyers, the consummate professional generalists, began to assume specialized tasks within firms.[5]

Along with this structural transformation came a shift in the content of legal work. With the expansion of railroads, oil companies, and other industrial giants, law firms faced a powerful demand for their services. They also faced clients who, for the first time, needed a substantially wider range of services and expertise than any single practitioner could provide. While lawyers at the start of the nineteenth century had also undoubtedly been involved in their clients' commercial and financial transactions, the size and complexity of these transactions increased rapidly towards the end of the century, as did the regulations seeking to guide and constrain them. These changes drove a need for larger firms.[6] Lawyers shifted their focus from courtroom argument to the preventive and more lucrative use of "office" law, and the sheer volume of services exploded. By the 1920s, increasing numbers of U.S. lawyers practiced in large firms and focussed on a new kind of law. No longer law-as-general-information, it was law-as-business, law as an increasingly collective enterprise.

Not all lawyers, of course, made the transition from individual to large-scale practice. And not all of the early law firms developed into professional behemoths. Those that did, however, evolved along very similar lines. Four changes, in particular, hastened the evolution of the most successful firms. First, their lawyers specialized, converting the traditional identity of generalist lawyers into narrowly focussed experts. Second, they realized—perhaps even created—advantages of size in an industry of intensely intellectual capital. Third, they redefined their core competencies, concentrating increasingly on the business of corporations. And finally, they developed an expanded capacity to manage information—not just through technical systems, but by managing the lawyers who directly possessed the most valuable stores of information. These four factors—specialization, size, a shifting focus, and the management of information—facilitated the invention of the modern law firm. Similar factors would again be critical in the next phase of evolution, from local firm to multinational practice.

The Growth of Global Law

During the 1940s and 1950s, U.S. law firms expanded in scale and scope —but not in geographical scope. They stayed remarkably fixed, to the point where many firms automatically identified themselves in part by their location: "the New York firm of Kelley Drye & Warren," for instance, or "the Washington, D.C., firm of Covington & Burling."

It wasn't until the 1960s that geographical expansion began in earnest. Initially, firms expanded to a small and select group of cities, often chosen for idiosyncratic reasons. They moved to Washington to be closer to the rulemakers and political power, and to New York to gain access to the dealmakers and financial elite. They moved to follow retiring clients[7] and to keep pace with the economic growth of sunbelt states and suburban industrial parks. Most of this expansion involved mergers rather than the establishment of new practices, and nearly all of it involved the existing large firms. By 1980, 87 of the largest 100 legal practices in the United States had branch offices and 24% of large-scale firms had a presence in three or more locations.[8]

International expansion occurred more slowly—and with good reason. From the outset, firms realized that expanding abroad was a far trickier enterprise than building either a corporate firm or a regional network. To do it well, firms would have to change not only how they ran their practices, but also, essentially, how they defined their professional competence.

The first complication lay with the basic contradiction of global law.[9] The assets U.S. law firms controlled—the information, expertise, and experience—were distinctly, completely, American. In the United States, these assets had considerable value, especially because law firms retained a virtual monopoly over them. Such assets were also useful to U. S. corporations abroad, insofar as they needed to stay within the confines of U.S. law and remain comfortable with the standards of commercial practice in their home market. But what use was knowledge of U.S. law for foreign clients in overseas markets? And what, if anything, could U.S. law firms hope to sell to these clients?

In grappling with these questions, law firms encountered a fundamental characteristic of information. To be of commercial value, information must be relevant to the purchaser. Enterprises buy the information they need and the services that help them fulfill their particular business functions. This specific demand provided the niche that law firms grew to fill. Over the course of their evolution, legal practices accumulated knowledge and expertise that business enterprises considered crucial to their own success. According to some analysts, in fact, the very prominence of U.S. law firms came from their ability to hoard valuable stores of information.[10] Only lawyers knew the intricacies of their legal specialty; only lawyers could argue before the court. So when individuals or corporations encountered these issues, they had few options but to call in the lawyers.[11] This necessity gave rise to the power and prestige of U.S. law firms. It also compelled U.S. corporations to take their lawyers with them as they expanded abroad: to write contracts that would stand in U.S. courts and arrange cross-border deals that maximized advantages under U.S. tax law. This necessity did not, however, extend to foreign corporations or individuals. On the contrary, insofar as law firms' knowledge rested on U.S. law, foreign entities—concerned presumably with their own laws—had no apparent need for the services of U.S. law firms. The firms controlled information-based assets that would appear to have little intrinsic value outside the borders of the United States.

There were also, ironically, basic issues of legality. Because law is so inherently political, nearly all countries have restricted the practice of law to their own nationals. Foreign lawyers, presumably unfamiliar with the intent of local law, have generally been forbidden from practice, or at least severely limited in terms of the services they could provide.[12] While the specific regulations have varied widely,[13] even the most open regimes have typically restricted foreign lawyers from practicing in at least five areas: litigation, domestic relations, the transfer of real property, wills and trusts, and advice on local law.[14] Many states have gone much further, confining foreign lawyers to advising on the laws of their home state, or subjecting them to rigorous, and often insurmountable, licensing requirements.

Not all the constraints have been regulatory. In most parts of the world, law has been (and remains) a business of relationships, where contacts are all-important and informal rules dictate the codes of conduct. In the clubby environment of client links, any foreign lawyer was, by definition, an outsider. And American lawyers, with their orientation towards corporate law and penchant for aggressive deal-making, have stood nearly as a separate breed—too greedy, too businesslike, and far too American. First in Europe, and then in Asia, local politicians have made clear their intent to resist, at least publicly, any encroachment by U.S. firms. "Their way of lobbying U.S. congressmen or senators cannot be of use here," declared a spokesman for the European Parliament. "It's really too pushy for Europeans."[15]

Much of this perceived "pushiness" came from the connection between American lawyers and American-style lobbying. While lobbying has been a private affair in most of the world. American lawyers have customarily approached lobbying as a public game of persuasion—a raucous and expansive game where fees and networking muscle can enter and affect the political arena. Even though many of the early entrants had not built their U.S. practice around lobbying, the perceived connection between American law firms and American lobbying marked them nonetheless. To those accustomed to a quieter world, this form of lawyering was a direct assault. To the Europeans, explained one Brussels insider, lobbying was a "nasty" word: "It means dirty money, dirty laundry, sex, all the stereotypes—persuasion beyond the point where you should be persuaded."[16] Similar sentiments have emerged in Japan, Eastern Europe, China, and Latin America—indeed, in nearly every foreign market that U.S. law firms have tried to enter.[17]

Even without such daunting legal and cultural resistance, investing abroad has always been an expensive and high-risk option for U.S. law firms. The price tag for an expatriate lawyer is relatively high, since it inevitably includes living quarters, school tuition, and the like. Because the success of a legal practice depends so critically on relationships, law firms felt they couldn't skimp on the niceties afforded their lawyers nor could they rotate them through short-term assignments. So each new overseas office generally meant finding a competent lawyer, ensuring that he or she had complete mastery of the language, and then shipping that lawyer off for an extended period of time. Again, the importance of relationships usually dictated that the new office be run by a partner, thus depriving the home office of one of its most valuable assets. Moving partners abroad also meant diluting the central relationship of the law practice—the partnership which, even in its most rigid incarnation, still ran by relatively informal, personal coordination. Thus, many partnerships decided early on not to go abroad, or at least to restrict their overseas activities to a handful of small offices. To most firms, quipped one New York partner, "Foreign had always meant New Jersey."[18]

Yet despite all these problems, U.S. law firms did go abroad. As of 1989, the 250 largest firms had established 180 overseas offices, up from 124 in 1985.[19] By 1991, that number hit 252.[20] By 1994, nearly 2000 U.S. lawyers were plying their trade in overseas offices, and U.S. exports of legal services hit $1.6 billion.[21] Most of these lawyers, and most of these exports, were concentrated in a few exceedingly large firms. In 1992, U.S. law firms accounted for seven of the ten largest firms in the world, and twenty-five of the forty largest. The remaining three of the ten largest firms were British, as were most of the remaining top forty.[22] Few firms from any other countries had achieved the global reach and prominence that increasingly defined the leading U.S. law practices and a handful of their British counterparts. By the mid-1990s, these firms could rightfully claim to have converted the parochial profession of law into a global business. And they dominated the field.

Such rapid and dramatic expansion raises three questions. First, why did U.S. and British firms move so decisively into the global marketplace, given the multiple obstacles and their own reluctance to do so? Second, what made some firms notably more successful than others in establishing and maintaining an overseas network? And third, what general lessons do the lawyers offer about global sales of professional services?

U.S. firms went abroad in two distinct waves. During the first wave, from 1965 to 1985, firms went solely to service their long-standing U.S. multinational clients. Minnesota's Oppenheimer, Wolff, for instance, went to Belgium in 1969 at the behest of Control Data Corporation;[23] Shearman & Sterling, New York-based advisor to Citibank, opened a Paris office in 1967 just as Citibank was completing a decade of unprecedented overseas expansion;[24] and Sullivan & Cromwell built a topnotch international practice almost entirely by following its premier client, Goldman Sachs, around the world.[25] Some of the U.S. firms that went abroad matured past their initial clients, building sizeable independent practices in their newfound locations. Most of them, though, did not, and left their overseas posts once their clients' work was completed.

During the second wave, from the early 1980s to the present, law firms increasingly ventured abroad on their own. San Francisco's Morrison & Foerster opened a Hong Kong practice in 1983, even though it had no existing clients in the region and no international presence aside from a London office; and Sidley & Austin of Chicago moved to Singapore in 1982 without a single client commitment.[26] For many of these later-movers, expansion into foreign markets was essentially a defensive move. Worried about losing core clientele, firms ventured abroad to ensure they could meet the full range of their clients' expanding needs. Anticipation about the arrival of the European Single Market drew firms in to Europe, even compelling some to inflate their size and areas of expertise by merging or affiliating rapidly with newfound European partners.[27] Once established in Europe, many U.S. firms then turned their sights on Asia—especially Japan, where a high yen and soaring economy seemed to harken opportunity.

The more firms went abroad, the more others itched to follow. This desire was also fueled by a need to track the mounting competition—to be "at the party," as one analyst described the trend.[28] Reportedly, many firms lost sizeable sums at their overseas offices, keeping them open simply in the hope of turning profits sometime in the future.

Eventually, being "at the party" became integral to staying competitive. When only a few firms were abroad, serving just their core customers, the rules of the competitive game were straightforward. The bulk of U.S. law firms handled their clients' overseas work from traditional home offices, and referred any substantial foreign business to foreign firms with which they maintained informal relations. These foreign firms, reciprocally, would refer U.S.-based work to their American colleagues. But as soon as U.S. firms started to hustle their own work overseas, they no longer needed foreign partners. Referrals dried up. This angered the foreign firms, which often stopped referring their domestic clients back to firms in the United States. And thus the competition accelerated.

Once a few firms established sizeable international practices, they achieved a critical mass that made them particularly attractive to multinational clients. Rather than asking multinationals to spread their legal advising among a number of far-flung firms, multinational law firms could offer coordinated and consolidated service.[29] To prevent their steadiest clients from being lured away, even staunchly domestic firms were compelled to compete, at least partially, in terms of geographic scope. Each firm that entered the fray upped the ante for the others.

Just as this rivalry was gaining momentum, a new form of competition entered the marketplace. The transition from communist rule in Eastern Europe had stimulated a tremendous demand for accountants: to value assets undergoing privatization and settle the accounts of firms accustomed to Soviet-style bookkeeping. Accordingly, the large U.S. accounting firms were among the first businesses to establish themselves in Eastern Europe. Once in place, they expanded their range of service offerings, using their relationships with both business and government clients to diversify into advising, consulting, and general deal brokering. Without the regulatory obstacles placed before lawyers, accounting firms were soon taking on the kind of work that law firms had come to see as their own. And the lawyers, mired in a U.S. recession, resented the trespass. So they began to compete directly against the international accounting firms, tracking them across Eastern Europe and the emerging markets of Asia.[30] If competition from other lawyers could drag U.S. law firms out of domestic complacency, competition from *accountants* virtually catapulted them into action. The threat inherent in the accountants' move was that, free from domestic licensing requirements, they might actually be able to do the lawyers' work, encroaching upon the hoard of information and expertise that was the legal profession's most valuable commodity.

Expanding Abroad: Drivers, Pitfalls, and the Factors of Success

Like the earlier transition from small-scale to larger-scale firms, the expansion into international practice forced major changes upon the legal profession. Some firms adapted rapidly, seizing the international marketplace as profitable new territory and a source of competitive advantage. Others disdained foreign practice completely. Still others ventured abroad and floundered.[31] Thus, the emergence of global law has not made all firms global, nor has it made all global firms successful.

What has made some law firms more successful than others in establishing and maintaining an overseas practice? One immediate observation is that country of origin matters greatly. As noted earlier, nearly all of the major international law firms are from the U.S. or Great Britain, countries that practice common, as opposed to civil, law. In explaining this phenomenon, lawyers occasionally refer to the inherent superiority of common law, or to the "better service" or "longer hours" that prevail among British and American firms. Despite the obvious parochialism, there may be an element of truth to such assertions. There are also, however, distinct historical reasons that explain why U.S. and British law firms were the first to expand internationally and why they still predominate in the international arena.[32]

Recall that in the United States, law went corporate at a relatively early stage. It happened towards the end of the nineteenth century, when the rapid industrialization and consolidation of American business compelled lawyers to focus their energies on the provision of business services and corporate law. Initially, this specialization mattered only in the context of the domestic market. But, in the aftermath of World War II, U.S. corporations and U.S. styles of corporate practice gradually spread across the international economy, thrusting U.S. law firms into an exceedingly strong position. Any foreign firm that wanted to do business in the United States or with a U.S. counterpart needed to have English language capabilities and lawyers familiar with U.S. law. So did any firm hoping to sell securities or obtain bank financing in the U.S. market. And so, eventually, did any firm trying to make, or avoid, a U.S.-style acquisition. Merely by virtue of specializing in corporate law, by developing a deep-seated expertise in financial and business transactions, and by speaking English, U.S. law firms stumbled upon a significant global advantage.

A similar dynamic prevailed in Great Britain, where London's position as a global financial center attracted foreign enterprises and drove them to arrange their transactions in accordance with local statutes. Accordingly, for both British and U.S. law firms, the early ascendance of large-scale capitalism, combined with the economic power of the state, created a global demand for the specific expertise of locally trained lawyers. This demand pulled U.S. and British lawyers into foreign markets and granted them a distinctive first-mover advantage. It also allowed these law firms to create a second key asset for overseas practice: size.

Like the multinational clients they followed abroad, U.S. and British law firms quickly gained a considerable advantage in foreign markets by virtue of their relative size. As Table 1 indicates, U.S. law firms with the largest overseas practices also generally rank among the largest law firms in the United States. Firms don't have to be international to be big players in their home markets, but they do, it appears, have to be large at home in order to build even a moderate overseas practice. A similar relationship prevails in the United Kingdom, where all nine firms that rank among the forty largest practices in the world also rank among the fifteen largest in England and Wales.[33] By itself, this correlation does not mean much, because firms with overseas branches will tend naturally to be larger than those with only a single-market focus. But a closer look reveals a competitive causality as well.

In foreign markets, law firms cannot offer a standard product. By law, or by custom, they cannot sell basic legal services. Given their relatively high overhead costs, they also cannot compete on price. So the only way for foreign firms

TABLE 1. U.S. Law Firms with Largest Overseas Practices[a]

Firm	Number of Overseas Lawyers[b]	Total Number of Lawyers	U.S. Rank by Gross Revenue[c]
Baker & McKenzie	1,377	1,897	2
White & Case	273	699	16
Coudert Brothers	148	348	62
Cleary Gottlieb	144	520	8
Jones, Day	82	1,204	3
Skadden, Arps	82	1,150	1
Shearman & Sterling	72	622	9
Davis Polk & Wardwell	46	515	10
Hogan & Hartson	46	459	47
Graham & James	44	406	74
Debevoise & Plimpton	41	368	24
Altheimer & Gray	40	209	n/a
Milbank, Tweed	38	350	31
Sullivan & Cromwell	35	440	4
LeBoeuf, Lamb	34	549	27
Rogers & Wells	34	440	34
Squire, Sanders & Dempsey	34	367	51
Willkie, Farr & Gallagher	32	400	29
Morgan, Lewis & Bockius	31	833	7
Weil, Gotshal & Manges	30	610	5

Notes: a. Size based on total number of lawyers in non-U.S. offices, 1996
b. From "The NLJ 250," *National Law Journal*, September 30, 1996
c. From "The AmLaw 100," *The American Lawyer* (July/August 1996)

to woo clients in distant markets is by offering differentiated services, information or advice unavailable from local sources. In foreign markets, law firms have to compete through specialization.

This is precisely what the successful U.S. and British firms did. They started with their obvious specialty, counseling foreign clients on the intricacies of U.S. law and corporate practice. Then they segued into increasingly specific areas of corporate strategy, advising clients on topics such as acquisitions, hostile takeovers, and debt restructuring—complicated corporate maneuvers that had all been developed first in the U.S. and British markets. As these sorts of deals proliferated across the international economy, U.S. and British firms found themselves with a natural, lucrative niche.

However, maintaining this niche entailed constantly refining and expanding their areas of specialization. Increasingly, clients who purchased business advice or strategic counsel wanted to rely on a single, central source of information rather than juggling across multiple sites. As a result of this logical preference, specialization compelled growth. Once firms achieved a certain critical mass, they could begin to amortize their fixed assets—their libraries, offices, information systems—more efficiently than their smaller competitors. They could also offer clients a more encompassing product. Small firms, almost by definition, did not have sufficient resources to craft and document all the materials demanded by a complex, fast-paced transaction. To handle these transactions in-house, firms had to be big, offering a well-stocked cluster of specialties, along with a critical mass of expertise in each field.[34] They then took their critical mass and marketed it explicitly, gaining business by promising clients an ability to deliver, as a partner at one firm put it, "everything you could want under one roof: seven relevant jurisdictions and three key languages."[35]

As these massive firms grew and prospered in the late 1980s, they discovered that size brought the benefit of reputation, as well, adding directly to firms' ability to generate new business. "Large size does not necessarily impart prestige," one lawyer noted, "but it comes very close."[36] In foreign markets, prospective clients often had no prior knowledge of individual law firms and had few ways of judging the quality of legal services. As risk-averse consumers, they were particularly prone to purchasing legal services only from well-known, big-name firms. Reputation was equated with competence and size became a measure of quality. Or as *Business Week* reported in 1986, "ever-growing corporate clients seem to feel more comfortable with big service firms and their 'brand-name' predictability."[37] Once again, this dynamic further accelerated the rise of the large, established firms.[38]

But size alone still could not rid U.S. or British firms of their outsider status. The only way around this obstacle was to go local, a strategy that most of the internationally successful U.S. and British firms began to adopt by the late 1980s. Rather than diverting and diluting competence at home, firms such as New York-based White & Case and London's Clifford Chance built competence abroad. They aggressively hired local lawyers in their target markets and constructed around them new networks of contacts and expertise—an acquired base that was critical to the law firms' success.[39] It gave them credibility in their target local markets and added new information to factor into their global networks. It also eased them around most regulatory and cultural barriers. By hiring locally trained attorneys, U.S. and British law firms could reap the advantages of both worlds, and then use these advantages to extend their global reach.

However, each foreign office drew energy away from the traditional locus of power. Each foreign hire diluted the firm's all-encompassing culture and raised questions of management and quality control. Firms were losing the ability to track and monitor their key asset: lawyers. To make global networks succeed, law firms needed to develop new modes of transnational management. The task of creating these modes has flustered even some of the powerhouse firms, leaving them either with too little control over their far-flung assets (and consequently with problems of reputation and quality) or with too much control (and a resulting lack of commitment from the overseas contingent). Intuitively, firms recognize the need to maintain sufficient control at the center without neglecting the needs and talents of the periphery. But calibrating the balance is difficult.[40]

Accordingly, many law firms have recently pulled back from the international arena, preferring to keep their partners in geographical proximity.[41] Firms reported being "burned" by their overseas efforts—of monies being spent without reward, of partnership squabbles, and of problems with recruiting and retaining sufficient foreign expertise. Others have seen their most cherished assets—partners with strong local expertise—depart abruptly, taking with them years of contacts and valuable experience. And still others have contented themselves with creeping warily along the international frontier, promoting the occasional foreign partner, but never going so far as to integrate their international activities into their core practice.

For those that have made the leap, however, the rewards appear substantial. As the data in Table 2 indicate, the U.S. firms with the largest overseas practices rank generally among the most successful legal practices. While the data by no means suggest that the most international firms are the most profitable (they're not), they do indicate that being international does not compromise profitability. Indeed, all of the largest international practices are profitable.[42] Their average profits per partner are 37% higher even than that of the AmLaw 100 (a ranking of the nation's 100 highest grossing law firms) and 9% higher than the average of the AmLaw 50.

In reviewing the international track record of these firms, the lessons for other lawyers seem obvious. To compete successfully in the international economy, law firms need to be big, well-known, locally specialized, and managed through a delicate balancing of global and local interests. Some firms may be able to depart from this model by developing distinctive competencies in highly specific areas (aviation or environmental law, for example), but for most firms in most foreign markets these four factors are critical to success.

The lawyers' experience tracks quite closely the experience of other information-based firms. None of the factors that drive success in the legal profession are specific to the legal field; all have been cited in other industries and different contexts. For a long time, lawyers have tended to see themselves as a profession, not a business. They have claimed idiosyncrasies and public responsibilities that put them above the commercial fray. And to some extent the distinctions are valid. Yet, once law firms begin to expand abroad, the parallels with other industries loom large. Multinational law firms are novel enterprises, but they are not unique. While lawyers may have followed down a familiar path in expanding abroad, their experience exemplifies the broader issues involved in selling intangible products across an international marketplace.

TABLE 2. Relative Position of U.S. Firms with Largest Overseas Practices

Firm[a]	Gross Revenue (000s)	Revenue per Lawyer	Profits per Partner
Baker & McKenzie	594,000	335,000	430,000
White & Case	246,500	445,000	565,000
Coudert Brothers	120,000	370,000	310,000
Cleary Gottlieb	286,000	650,000	910,000
Jones, Day	400,000	395,000	425,000
Skadden, Arps	635,000	635,000	885,000
Shearman & Sterling	283,500	520,000	595,000
Davis Polk & Wardwell	282,000	740,000	975,000
Hogan & Hartson	140,000	335,000	465,000
Graham & James	100,000	350,000	260,000
Debevoise & Plimpton	196,500	590,000	890,000
Altheimer & Gray	n/a	n/a	n/a
Milbank, Tweed	173,000	530,000	590,000
Sullivan & Cromwell	317,500	820,000	1,310,000
LeBoeuf, Lamb	180,000	360,000	360,000
Rogers & Wells	164,000	450,000	475,000
Squire, Sanders & Dempsey	132,000	375,000	330,000
Willkie, Farr & Gallagher	176,000	515,000	760,000
Morgan, Lewis & Bockius	296,000	405,000	405,000
Weil, Gotshal & Manges	305,500	530,000	705,000
Average	**264,610**	**492,110**	**612,890**
AmLaw 100 Average	**162,140**	**426,000**	**446,000**
AmLaw 50 Average[b]	**222,950**	**475,700**	**560,600**

Notes: a. The firms listed here are those with the largest overseas practices, as presented in Table 2.

b. "AmLaw 50" refers to the top 50 firms of the AmLaw 100 (measured by gross revenue).

Source: Table 2: "The AmLaw 100 Firms," *The American Lawyer* (July/August 1996). All data are the most recently available for completed fiscal years; in some cases, *American Lawyer* may estimate.

Selling the Unseen: Lawyers and Their Peers

What general lessons can be learned from law firms about global sales of professional services? Law, as a business, bears a strong resemblance to advertising, management consulting, and accounting—to any business where the product being sold consists primarily of knowledge and the production process occurs largely in people's minds[43] Where these conditions exist, parallels from the legal field are most directly applicable.

TABLE 3. Largest U.S. Law Firms, 1985 and 1995

Firm	Revenues 1985 (US$000s)	Lawyers 1985	Revenues 1995 (US$000s)	Lawyers 1995
Skadden, Arps	169,000	526	635,000	1,001
Baker & McKenzie	128,000	751	594,200	1,776
Jones, Day	108,000	435	400,000	1,014
Sullivan & Cromwell	94,000	251	317,500	387
Weil, Gotshal & Manges	90,000	286	305,500	575
Latham & Watkins	89,000	302	299,500	575
Morgan, Lewis & Bockius	91,000	388	296,000	732
Cleary, Gottlieb	81,000	229	286,000	441
Shearman & Sterling	117,000	378	283,500	546
Davis Polk & Wardwell	118,000	278	282,000	381

Source: "The AmLaw 100," *The American Lawyer* (July/August 1996) and (July/August 1986)

The Value of Size

In the international marketplace, size has value. While many lawyers remain sole practitioners and others have grouped themselves into small but highly specialized boutique firms, nearly all of the most profitable and highly regarded firms are big. Indeed, as Table 3 indicates, the top tier law firms have grown substantially larger over the past decades and show no signs of slowing the pace of their growth in the near future.[44] In the process, these expanding firms are causing smaller firms to either consolidate or settle into highly specialized niches. As a result, one partner noted, "there is no middle class left."[45]

This combination of specialization and enlargement is neither a coincidence nor a passing phenomenon. It is a strategic response. While small firms can develop deep and specific expertise, only big firms can bundle separate areas of expertise into an integrated analysis. Only big firms can be simultaneously specific and broad, acquiring a "credible depth" to handle complex transactions.[46] Only big firms have the internal capacity, as one partner explained, to "generate vast amounts of paper and distribute it to the four corners of the world in 48 hours."[47] Big firms can develop economies of information that enable them to dominate the market for its sale. They can hoard expertise, hone it, and offer salary inducements that smaller competitors cannot possibly match. They can produce and distribute (often for free) publications that simultaneously disseminate information and market their own services. They can afford to "buy off" the time of some of their most highly valued assets, granting them sabbaticals or publishing venues or other means of replenishing and displaying their own stores of knowledge. Big firms are also in a better position to establish and maintain increasingly sophisticated networks for managing information flows across geographically dispersed offices.

This power of size is apparent across the information-intensive industries. Not only have law firms grown bigger over the past two decades, but so have accounting firms, media firms, advertising agencies, and banks.[48] As the pool of information expands and flows across international borders, firms that process information are intuitively recognizing the benefits of size. To compete, they must be all things to all customers. Yet the markets they serve are different, and their products, unlike manufactured goods, can be neither stockpiled nor standardized. Thus information-based firms in a global economy must be able to adapt and change along with their context. To do this well, they have to be big.

Nevertheless, the correlation between size and success is neither infinite nor perfect. There are some small firms, such as New York's Haight, Gardner, Poor, and Havens (specializing in maritime law), that have done extremely well in overseas markets. And there are some large firms, such as the venerable Coudert Brothers, that appear to have become victims of their own size, entangled by over-heavy management processes. Just where the threshold lies is difficult to determine with any clarity. But both the available data and anecdotal testimony strongly suggest that, in the international marketplace, size provides information-based firms with a considerable competitive advantage. Despite the internal tensions to which it is inherently prone, the large global law firm, as one partner asserted, "is clearly the wave of the future."[49]

The Value of Reputation

For information-based firms, the importance of reputation stems from the basic inability of customers to measure the quality of the information they receive: in the short term, a corporation cannot tell if it received the best contract, the optimal tax strategy, or the most suitable acquisition target. It cannot objectively judge either value or competence. Thus, reputation serves as an all-important proxy. Firms buy the advice or expertise of a particular provider because others have spoken well of its talents, or because its name alone connotes commercial heft.[50] Because objective or mechanical evaluations of information-based products are difficult to construct, customers are forced instead to rely on the market's perception of quality.[51]

In foreign markets, particularly less developed ones, reputation carries even greater weight. With only scant knowledge of overseas providers, clients are wary of "bad" purchases and more anxious to buy a well-known, brand-name commodity. They want, as one partner described it, to procure a "Good Housekeeping Seal of Approval" for risky deals and, occasionally, to boost their own reputation by allying themselves with a globally recognized firm.[52] Consequently, well-known information providers can expand their global sales simply by being well-known information providers. And smaller firms, scrambling to establish their name and credentials, are likely to be left at a serious disadvantage. This preference for reputation clearly marks the record of law firms' expansion: it is also evident in accounting, management consulting, health care services, and investment banking.[53]

The Importance of "Walking Assets"

One of the characteristics that distinguishes information-based firms from manufacturing firms is their central reliance on people. This isn't to imply that manufacturing firms do not need or value their employees, but only that the relationship for information-based firms is tighter and more critical. The difference lies with the relative weight of human, versus physical or financial, capital. In most information-based industries, the "factory" is little more than an agglomeration of people. There may be some real estate and computer hardware, but the firm's assets reside principally in its people. This concentration creates distinctive challenges for information-based firms. To compete, firms must acquire the most profitable and productive assets, which means hiring and retaining the most talented people. It also means training these people in particular specialties, enabling them to develop the expertise so critical to the firm's success, and then integrating them into a system for coordinating and consolidating their specific knowledge. Once they become part of this system, the firm's "walking assets" are highly valuable and difficult to replace. Retention is thus key, particularly in foreign markets where local knowledge is liable to be concentrated in just a handful of people. With walking assets in foreign markets, a central objective for information-based firms is to ensure that these assets don't walk away from the firm.

The Importance of Balancing Global and Local Interests

Many scholars have demonstrated that the key to success for multinational firms is to "think globally and act locally."[54] This relationship applies with particular force to information-based firms. Successful overseas legal practices such as White & Case and Clifford Chance "went native" from the start, either hiring local attorneys or sending partners with a personal interest in the region and the intention of spending their careers there.[55] They learned, as one partner wryly noted, that "you can't just parachute people in from New York."[56]

However, managing across a web of local outposts is a complicated and delicate task. It entails providing autonomy to local offices while maintaining control at the center—a feat far easier said than done. One firm, the Chicago-based behemoth of Baker & McKenzie, claims to have created a management structure that achieves this ideal. Nearly all of the firm's lawyers work in their native country and receive compensation based on local returns. Local offices retain full autonomy over their own operations, functioning effectively as independent firms and adapting readily to local custom. All major decisions, however, are pushed up to a cross-national policy committee, manned by partners from across the practice.[57] According to Baker's management, this structure enables the firm to operate "as a global legal business."[58] But outsiders are far less complimentary. They deride Baker as a franchise rather than a firm, the

"McLaw" of the legal world. More importantly, both competitors and clients repeatedly cite quality control as a problem for Baker's practice. With such far-flung offices and disparate styles of lawyering, the firm appears to suffer a serious lack of consistency, one that keeps Baker & McKenzie, despite its enviable global reach, from being considered among the top tier of international firms. Other firms have had even greater difficulties. Recently, for example, Coudert Brothers, one of the first firms to establish a transnational practice, has vacillated wildly between having too much and too little control at the center.[59] In the course of these vacillations, attorneys have wearied of the practice and left, taking with them considerable chunks of the firm's reputation.

The extrapolation from this case is stark. To succeed in global markets, information-based firms need simultaneously to be big and highly specialized, global and local. They must attract the best talent, retain it, and then use this talent to establish an international network of contacts and a global reputation. Then they have to manage this huge and fragile enterprise, composed of rather huge and fragile egos. If the firm can not keep these essential walking assets happy, then they will, in fact, walk. The firm will consequently lose both the specific expertise vested in these people as well as their personal cluster of contacts. The firm's reputation, and its business, will suffer.

The experience of the legal profession thus suggests that selling information in the international marketplace entails considerable risk and requirements. It demands a scale of activity, a level of specialization, and an organizational structure far beyond the dictates of a single domestic market. It means managing information and people to an extent unheard of in many information-based firms, especially those that see themselves as professional or creative organizations rather than mere commercial enterprises. Yet the lawyers' experience also suggests that these requirements can be met. Not by all firms, and not easily, but the advantages of size and reputation mean that those that adapt most quickly stand to gain extremely lucrative rewards.

Just how to make the appropriate adaptations is a strategic decision that can only be made on a case-by-case and firm-by-firm basis. But in reviewing the options and charting a course, firms in other information-based industries would do well to watch—and learn from—the lawyers.

Notes

1. There is, of course, a vast literature that defines the differences between goods and services, and the further distinctions of information-based services. See, for instance, T.P. Hill, "On Goods and Services," *The Review of Income and Wealth*, 23 (December 1977): 315-331; Geza Feketekuty, *International Trade in Services* (Cambridge, MA: Ballinger Publishing, 1988); Jagdish Bhagwati, "International Trade in Services and Its Relevance for Economic Development," in Orio Giarini, ed., *The Emerging Service Economy* (New York, NY: Pergamon Press, 1987); Gary Sampson and Richard Snape, "Identifying the Issues in Trade in Services," *The World Economy*, 8 (June 1985): 171-182.

2. In 1743, for example, there were only eight lawyers in New York City. See Roscoe Pound, *The Lawyer from Antiquity to Modern Times* (St. Paul, MN: West Publishing Co., 1953), p. 141. Some of Pound's most provocative insights come from John Adams's diary. See John Adams, *The Works of John Adams* (Boston, MA: Little, Brown, 1850—1856).
3. Alexis de Tocqueville, cited in James W. Hurst, *The Growth of American Law* (Boston, MA: Little, Brown, 1950), p. 251.
4. See for instance Marc Galanter and Thomas Palay, *Tournament of Lawyers: The Transformation of the Big Law Firm* (Chicago, IL: University of Chicago Press, 1991), pp. 7-8. Their work builds upon Theron Strong, *Landmarks of a Lawyer's Lifetime* (New York, NY: Dodd, Mead, 1914); Charles Chamberlayne, "The Soul of the Profession," *Green Bag*, 18 (1906): 396-401. See also Wayne K. Hobson, *The American Legal Profession and the Organizational Society 1890-1930* (New York, NY: Garland Publishing, Inc., 1986), pp. 150-152; Hurst (1950), op. cit., p. 308; Thomas Paul Pinansky, "The Emergence of Law Firms in the American Legal Profession," *University of Arkansas at Little Rock Law Journal*, 9 (1986-87): 606-608 and 612-614.
5. Two other structural changes occurred in the legal profession during this same phase: the growing importance of law schools and the growing reach and role of bar associations. For more on these linked developments, see Wayne K. Hobson, *The American Legal Profession and the Organizational Society 1890-1930* (New York, NY: Garland Publishing, Inc., 1986), p. 76. For a discussion, see Pinansky, op. cit., pp. 593-640.
6. See Pinansky, op. cit., p. 611.
7. Between 1974 and 1980, 29 New York firms opened branch offices in Florida. Richard Abel, *American Lawyers* (New York, NY: Oxford University Press, 1989), p. 189.
8. Galanter and Palay (1991), p. 47.
9. Throughout this article, I use the terms "global" and "international" interchangeably, despite several excellent descriptions of the implied differences between these two terms. See for instance, Christopher A. Bartlett and Sumantra Ghoshal, *Managing Across Borders: The Transnational Solution* (Boston, MA: Harvard Business School Press, 1989); Yves L. Doz and C.K. Prahalad, *The Multinational Mission: Balancing Local Demands and Global Vision* (New York, NY: Free Press, 1989). My use of "global" or "international" law also refers to the overseas activities of law firms, rather than to the formal body of public international law, which concerns itself primarily with relations between sovereign states. To practitioners, by contrast, as Flood notes, international law means advising on transactions or litigation involving more than one jurisdiction. See John Flood, "Megalawyering in the Global Order: The Cultural, Social and Economic Transformation of Global Legal Practice," *International Journal of the Legal Profession*, 3 (March 1996): 189.
10. See for instance, Robert L. Nelson, *Partners with Power: The Social Transformation of the Large Law Firm* (Berkeley, CA: University of California Press, 1988), p. 266; Pinansky, op. cit., pp. 634-635.
11. One obvious alternative for corporations is to create inhouse legal staffs, thereby reducing both financial and transaction costs.
12. For a discussion of current developments concerning the future of "foreign legal consultants," see Sidney M. Cone III, "FLCs face licencing hurdles in host countries," *The National Law Journal*, October 7, 1996, p. C2.
13. France, for instance, allowed only members of the legal and judicial professions *in* France to give legal advice. See EU Schedule on Professional Services; cited in Kenneth S. Kilimnik, "Lawyers Abroad: New Rules for Practice in a Global Economy," *Dickinson Journal of International Law*, 12 (Winter 1994): 269-325. In Japan, foreign lawyers were allowed to practice, but only on issues that concerned the

law of their home state, and only after receiving approval from the Japanese bar association. [Special Measures Law Concerning the Handling of Legal Business by Foreign Lawyers, art. 10(3).] See also generally Amelia Porges, "Introduction to Japan: Law and Ministerial Order Related to the Handling of Legal Business by Foreign Lawyers," *International Legal Materials*, 26 (1987): 881-920. Foreign law offices were also not permitted to hire Japanese attorneys. In 1987, Japan liberalized its laws somewhat, permitting foreign lawyers to petition to become "foreign legal consultants" (and thereby practice home-jurisdiction law) under certain conditions. Until the spring of 1996, Hong Kong permitted foreign lawyers to practice, but only if they confined their advice to the laws of their home states. See Janice Fuhrman, "Hong Kong is Adjusting to the '1997 Jitters'," *The National Law Journal*, July 11, 1983, p. 26. Hong Kong law also prohibited local Hong Kong solicitors from employing any employee of a foreign law firm, or even from sharing office space with them. The Hong Kong regulations have been particularly despised by U.S. firms, because they treat British law firms as local, rather than foreign, entities. See Joan M. Cheever, "U.S.-Hong Kong Deals Shaky?" *The National Law Journal*, August 20, 1990, p. 3; Simon Ip, "Legal Skills Cannot Cross National Boundaries," *Far Eastern Economic Review*, November 17, 1988, pp. 44-46.

14. See Richard L. Abel, "Symposium: The Future of the Legal Profession: Transnational Law Practice," *Case Western Reserve Law Review*, 49 (1994): 737-870.
15. Quoted in Sheila Kaplan, "Backlash in Brussels," *Legal Times*, December 3, 1990, Report from Europe, p. 1.
16. Ibid.
17. See, for instance, Steve Albert, "Looking South of the Border," *The Recorder*, May 28, 1991, p. 12; Joan M. Cheever, "The Legal Community Withdraws its Welcome in Hong Kong," *The National Law Journal*, January 9, 1989, p. 22.
18. Quoted in Benjamin Weiser, "204 Miles and a World Apart," *Washington Post*, February 10, 1994, p. D11.
19. Sherry Sontag, "Opening Offices Overseas: Does the Payoff Warrant the Huge Expense?" *National Law Journal*, February 12, 1990, p. 1.
20. Calculated from Abel (1994), op. cit., Table 23, p. 859.
21. The export figure includes only payments by foreign-based sources to U.S.-based firms, thus capturing only a portion of total U.S. legal revenues sourced abroad. See John E. Morris, "Capitalizing on Global Capitalism," *The American Lawyer* (April 1996), p. 5. As Abel notes, the overseas activities of U.S. lawyers still constitute only a tiny fraction of the total revenues for the profession; but the growth rate and absolute number are remarkable. See Abel (1994), op. cit., p. 738.
22. See Abel (1994), op. cit., p. 741; and Robert Budden, "The IFLR Top 40: The World's Largest Law Firms," *International Financial Law Review*, October, 1992, p. 13. The remaining firms in the top 40 include four Australian and one Canadian practice.
23. Deborah Graham, "Two Decades See Growth in Overseas Branching," *Legal Times*, May 28, 1984, p. 9.
24. Sometime later, Shearman also followed Citibank to Latin America to advise the bank on its massive restructuring of country debt. Subsequently, the firm's Latin American practice took off. See Sam Adler, "Hola Latinoamerica," *Manhattan Lawyer* (September 1991), p. 22.
25. See Karen Dillon, "Brand Names at the Brink," *The American Lawyer* (May 1995), p. 5.
26. Graham (1984), op. cit., p. 9.
27. One analyst observing this movement attributes it to the Thomas Theorem: that "trends become real, because people believe them to be real." See Erhard Blankenburg, "The Thomas Theorem of mega-lawyering," in Philip S.C. Lewis, ed., *Law*

and Technology in the Pacific Community (Boulder, CO: Westview Press, 1994), p. 28. See also Timothy Harper, "Going Global: Big Law Firms Expand Overseas," *ABA Journal* (September 1989), pp. 68-72.

28. Abel (1994), op. cit., p. 741.
29. As Galanter and Palay note, "one stop shopping" benefits the corporate buyer by reducing both transaction costs and search costs. See Galanter and Palay (1991), op. cit., p. 50.
30. For more on the emerging market squabbles between accountants and lawyers, see Sheila Kaplan, "Eastern Europe: Dealmakers Rush to a New Frontier," *The American Lawyer* (April 1991), Corporate Scorecard 1990, p. 6; Chris Klein, "Gold Rush, Thin Stakes," *The National Law Journal*, August 12, 1996, p. A1.
31. In at least some cases, however, poor performance of foreign offices turns out to be a challenge, not a death blow. See Emily Barker, "Making Milbank Over," *The American Lawyer* (October 1996), p. 43.
32. See "Red Tape Around the World," *The Economist*, November 23, 1996, p. 77.
33. Drawn from data presented in John Flood, "The Cultures of Globalization: Professional Restructuring for the International Market," in Y. Dezalay and D. Sugarman, eds., *Professional Competition and Professional Power: Lawyers, Accountants and the Social Construction of Markets* (London: Routledge, 1995), Table 6.1, p. 151; and "The IFLR Top 40: The World's Largest Law Firms," *International Financial Law Review* (October 1992), p. 13.
34. For more on this phenomenon, see Paul Hoffman, *Lions of the Eighties* (New York, NY: Doubleday, 1982), pp. 226-233. For a discussion of how the trend toward "boutique-ing" is liable to affect the future of the legal practice, see Steven Brill, "The Law Business in the Year 2000," *The American Lawyer*, June 1989, p. 5.
35. Interview, London, October 1996
36. Mark Rollinson, "Small Company, Big Law Firm," *Harvard Business Review* (November/December 1985), p. 7. For more on the attributes of size, see also the discussion in Galanter and Palay (1991), op. cit., p. 50.
37. William B. Glaberson, "Megafirms are taking over corporate law," *Business Week*, November 17, 1986, p. 104. For an expanded discussion of the factors behind law firm growth, see Nelson (1988), op. cit., pp. 62-63; and Arleen Leibowitz and Robert Tollison, "Free Riding, Shirking, and Team Production in Legal Partnerships," *Economic Inquiry*, 18 (1980): 380-394.
38. It also apparently increased the relative share of profits earned by the top tier firms. See Brill (1989), op. cit.
39. The importance of hiring local attorneys in order to build relationships and generate business is a recurrent theme. See for instance Ed Finkel, "Altheimer Expands Boundaries Again with Move to Turkey," *Chicago Lawyer* (June 1994), p. 65; David Trubek et al., "Global Restructuring and the Law: Studies of the Internationalization of Legal Fields and the Creation of Transnational Arenas," *Case Western Reserve Law Review*, 44 (Winter 1994): 407-498.
40. See, for instance, Michael Oppenhoff, "Organisation of International Legal Services," International Bar Association, Section on Business Law Seminar, Amsterdam, March 14, 1995.
41. For an interesting support of this position, see John Merrow, Senior Partner of Sullivan & Cromwell, transcript of presentation to the American Bar Association Annual Meeting, New Orleans, August 1994.
42. Altheimer & Gray, which is by far the smallest U.S. firm with a major overseas presence, is the only one that does not appear as an Am Law 100 firm. Since the Am Law 100 is calculated according to gross revenue, Altheimer & Gray's failure to place in this ranking is not necessarily consequential.

43. Software, publishing, and media have some, but not all, of the same characteristics. For a related discussion of the problems that surround the electronic sale of information, see Debora Spar and Jeffrey J. Bussgang, "Ruling the Net," *Harvard Business Review* (May/June 1996), pp. 125-133.
44. This point is also confirmed in James F. Fitzpatrick, "Legal Future Shock: The Role of Large Law Firms by the End of the Century," *Indiana Law Journal*, 64 (Summer 1989): 461-471.
45. Personal interview, Washington, D.C., September 1996.
46. The term was suggested during a personal interview, New York, NY, October 1996. For more on this point, see also Saundra Torry and B.H. Lawrence, "D.C. Area Law Firms Join Wave of Mergers," *Washington Post*, April 10, 1989, p. E1; Joseph Flom, "Is the Multicity Mega-Firm a Necessity or a Pipe Dream?" panel discussion, reported in *American Lawyer*, under the title "Getting Big: Is Growth Unavoidable?" (December 1984), pp. 3-15 of supplement.
47. Personal interview, New York, NY, September 1996.
48. In a a related discussion, DeAngelo argues that big accounting firms are inherently better than small ones, because they are less reliant on any single customer, and thus more likely to produce objective reports. See Linda A. DeAngelo, "Auditor Size and Audit Quality," *Journal of Accounting and Economics*, 3 (1981): 183-199. For a description of how the investment management industry is also splitting into two tiers of very large and very small firms, see Mark P. Hurley et al., "The Coming Evolution of the Investment Management Industry: Opportunities and Strategies," Goldman Sachs & Co., Investment Management Industry Group, October 1995.
49. Interview, London, October 1996. This comment was made repeatedly in my interviews.
50. Gilson and Mnookin argue that large firms can be seen in fact as a quality assurance mechanism. See Ronald J. Gilson and Robert H. Mnookin, "Sharing among the Human Capitalists: An Economic Inquiry into the Corporate Law Firm and How Partners Split Profits," *Stanford Law Review*, 37 (January 1985): 313-397.
51. Or as Lovelock and Yip note, "[t]he uncertainty engendered by intangibility requires strong branding to offset it." Christopher H. Lovelock and George S. Yip, "Developing Global Strategies for Service Businesses," *California Management Review*, 38/2 (Winter 1996): 64-86.
52. Personal interview, New York, NY, September 1996.
53. See, for instance, Feketekuty (1988), op. cit., pp. 48-50.
54. See Christopher A. Bartlett and Sumantra Ghoshal, "What Is a Global Manager?" *Harvard Business Review*, 7 (September/October 1992): 124-132; Christopher A. Bartlett and Sumantra Ghoshal, *Managing Across Borders* (Boston, MA: Harvard Business School Press, 1989); Yves L. Doz and C. K. Prahalad, *Multinational Mission: Balancing Local Demands and Global Vision* (New York, NY: The Free Press, 1987).
55. This was confirmed to me repeatedly in interviews conducted with law firms in New York, Washington, D.C., and London in the fall of 1996.
56. Personal interview, New York, NY, September 1996.
57. Major decisions include those that involve an investment of more than $1 million or a firm commitment of more than five years. See Deborah Graham, "Baker & McKenzie Streamlines Democratically," *Legal Times*, February 25, 1985, p. 12. For more on the practice and structure of Baker & McKenzie, see Barnaby J. Feder, "The Unorthodox Behemoth of Law Firms," *New York Times*, March 14, 1993, Section 3, p. 1; James B. Stewart, Jr., "Inside the World's Largest Law Firm: Baker & McKenzie," *The American Lawyer*, January 1980, p. 32; Mark Stevens, "Chicken

McLaw: The Evolution of Baker & McKenzie," *Crain's Chicago Business*, November 24, 1986, p. 25.

8. Donna Gill, "Baker's Unique Niche," *Chicago Lawyer* (January 1992), p. 1.
9. See Nicholas Varchaver, "The Culture of Chaos," *American Lawyer* (December 1993), p. 52.

Stopford & Strange,
RIVAL STATES, RIVAL FIRMS,
Cambridge University Press, 1991

1 THE NEW DIPLOMACY

The upheavals of the international political economy during the last decade have altered, irreversibly we believe, the relationships among states and multinational enterprises. Growing interdependence – that much abused word – now means that the rivalry between states and the rivalry between firms for a secure place in the world economy has become much fiercer, far more intense. As a result, firms have become more involved with governments and governments have come to recognise their increased dependence on the scarce resources controlled by firms. This mutual interdependence of states and firms throughout the world is the subject of this book; even though the detailed material is drawn from just three countries – Brazil, Kenya, Malaysia – we believe it raises new and universal questions just as relevant in Eastern Europe, the Soviet Union or China as in the third world.

We start our questioning with six general propositions. The first is that states are now competing more for the means to create wealth within their territory than for power over more territory. Where they used to compete for power as a means to wealth, they now compete more for wealth as a means to power – but more for the power to maintain internal order and social cohesion than for the power to conduct foreign conquest or to defend themselves against attack.[1] The implication is that national choices of industrial policy and efficiency in economic management are beginning to override choices of foreign or defence policy as the primary influences on how resources are allocated.

The second is that the emergence of new forms of global competition among firms also affects how states compete for wealth. As firms harness the power of new technology to create systems of activity linked directly across borders, so they increasingly concentrate on those territories offering the greatest potential for recovering their investments. Moreover, in a growing number of key sectors, the basis of competition is shifting to emphasise product quality, not just

WHARTON REPROGRAPHICS

981
741

costs. Attractive sites for new investment are increasingly those supplying skilled workers and efficient infrastructures. These new demands from firms affect how governments allocate resources to attract wealth-generating investment.

The third is that small, poor countries face increased barriers to entry in industries most subject to global forces of competition. They must look to their investments in skills as a primary means of hooking into the growing international systems and avoiding constant relegation to the periphery of investors' concerns.

The fourth is that these changes have added two new dimensions to diplomacy. No longer do states merely negotiate among themselves; they now must also negotiate – if not as supplicants then certainly as suitors seeking a marriage settlement – with foreign firms. Furthermore, multinational firms themselves are increasingly having to become more statesmanlike as they seek corporate alliances, permanent, partial or temporary, to enhance their combined capacities to compete with others for world market shares. The interaction of all three dimensions, in 'triangular diplomacy', calls for new skills in management and government that challenge the old order.

The fifth is that these new dimensions have multiplied the number of possible policy options for governments and for firms, and thus have greatly complicated the problems for both of managing multiple agendas. The administrative capacity of both has now become an important determinant of who can gain most from the changes in the world economy.

The sixth is that all of these shifts have acted to increase the volatility of change and the divergence of outcomes of the new diplomacy. Many developing countries are poorly placed to respond effectively, *not* for reasons of lack of factor-cost advantages, but because of deep-seated internal obstacles. These are born of traditional attitudes, political structures and often a lack of political will to confront and resolve inherent dilemmas of choice: policies aimed at enhancing internal welfare seldom sit comfortably with those designed to enhance the efficiency needed to compete in world markets. Development thus becomes a function of nations' abilities to link and control their economic affairs co-operatively with others: policies of autarky are increasingly ineffective.

These propositions all suggest reasons why policy must become more outward-looking if states are to find a place in the sun. But *how* states both resolve increasingly intractable dilemmas and implement policy becomes a critical determinant of success. For many, the prevailing attitude expressed in more liberal attitudes towards the multinationals seems to be, perhaps grudgingly, that of saying, 'if you cannot beat them, then join them'. Yet, at the same time, many are also saying that the struggle for self-reliance must go on. There are, thus, countervailing forces at work. The creation of fruitful partnerships in production will be, we believe, a fitful process and suffer many setbacks in the years ahead. But the direction of progress seems clearly charted, even if the details are blurred.

Outline for the book

In this introductory chapter, we summarise the general line of argument to amplify these initial propositions. In chapter 2, we elaborate our contention that global structural changes – in finance, technology, knowledge and politics – have exerted a dominant influence on the behaviour of governments and firms and pushed multinational enterprises more centre-stage in the evolving international political economy. In chapter 3, we examine how structures of global competition among firms have been affected by these shifts and why they have been evolving at different rates and in different forms. We argue that governments need to understand more fully that differences in firms' abilities to master change are more important than are differences among industries. Only by looking at the relative strength of firms in an industry can the basis of lasting bargains be discerned.

The next two chapters focus on the nitty-gritty detail of action and response at local levels. In chapter 4, we examine how our three governments have formulated national and sectoral policy and attempted to resolve the dilemmas inherent in their dealings with foreign firms and with international institutions such as the World Bank, the IMF and the web of treaty organisations such as the GATT. The need for control – to ensure that firms do not cheat, as well as to ensure order in managing economic and social transformation – can conflict with the promotion of entrepreneurship. The divergence of policies and outcomes in these countries seems to us especially striking and not susceptible to interpretation by any single model of bargaining power. Chapter 5 examines the impact of the global strategies of firms on their policies in developing countries. In the specific relationships between firms and governments, how the policies of each affect the responses of the other, whether positively or negatively, is illustrated. From this it will be clear that firms seek governments that provide a stable, temperate environment, while governments look for firms that will be good citizens, productive, expanding and loyal. Chapter 6, written by John Henley, examines the social dimensions of the government–corporate relationship. Working with the grain of what is socially and politically acceptable in each country is essential if policies are to be stable and to have a chance of surviving external shocks.

In the concluding chapter we return to our six propositions and amplify them in the light of the evidence presented. Here, we expand on the notions of the new dimensions to diplomacy and examine the implications of the additive and growing matrix of agendas that determine the interaction between firms and governments. We aim to persuade our readers that a few simple models can cut through much of the complexity by linking together agendas previously kept separate. The new questions thus provoked can help to identify and explain future sources of opportunity and risk for both parties.

THE CHANGING GAME

Our propositions suggest that structural change in the international political economy has altered the nature of the game by affecting the actions and responses among firms and states. What we mean by structural change will be explained in some detail in the next chapter. But for the moment we would point to some of the more visible outcomes of greater volatility in a world of more mobile financial resources and faster technical change.

A brief catalogue of some of the major sources of dynamism in world markets illustrates the difficulties of forecasting when determining appropriate policy for developing countries with differing resources. For example, modern designs and technology allow firms to save on materials. By 1984, Japanese firms had reduced by 40 per cent the raw materials needed for each unit of industrial production in 1973 (Drucker, 1986). Such a massive reduction directly affects the commodity-exporting countries: much of the persistent weakness in the prices for many commodities seems to have as much to do with this effect as with the production cycle. Simultaneously, the costs of labour as a proportion of total costs have been falling in all developed countries. In many fabricating industries, they have fallen from 25 per cent to under 10 per cent within the last decade (Ohmae, 1985); in TV, they fell to 5 per cent in Japan by 1980 and have fallen further since. Similar falls are endemic in the information-technology industry.[2] Developing countries that have based their export strategies on low labour costs now find those advantages to be eroding. Indeed, there have been some instances where investors have pulled back their labour-intensive operations from developing to developed countries (UNCTAD, 1989).

The real costs of transport and communication have also been falling. That firms have taken advantage of the reduced costs of running a far-flung corporate empire and altered the balance between trade and investment in their strategies is shown, for example, by the fact that the value of air freight across the Pacific overtook passenger revenues in 1986. Moreover, reduced communication costs allow new *systems* of global information management to become feasible. If the 1980s were the decade of the personal computer – 110 million sold worldwide – the 1990s look set to be the decade of telecoms and other forms of electronic communication. Advantage will shift to those who can take advantage of collapsing time scales and harness the value of the signals received by innovative action.

In some sectors, these shifts have stimulated trade in intermediate goods; in others, more trade in finished goods and less in the intermediates previously produced in developing countries. Added to these changes are the increasing costs of R & D, the shortening life for many products, the new possibilities for producing variety at lower cost, and the increase in the risks for many new product ventures. These and other factors have combined to create novel, and often unpredicted, divisions of labour within and among developed and developing countries.[3] They have also changed the structure of entire industries. In the pursuit of ever lower costs and greater shares of an increasingly open world market, many firms have been propelled into taking on the challenges of what has become known as global competition.

Shifting opinion

As these changes have unfolded, the tone and substance of the debates about the role of multinational companies in developing countries have been revised. It is hard to remember that only a decade ago politicians in third world countries were almost unanimous in their castigation and condemnation of foreign companies, and that the heads of many of these companies, especially the American and European ones, held inflexible views on how they would operate in a developing country. In place of the old bitterness, bigotry and mutual incomprehension,[4] we find a new pragmatism in their mutual attitude.

For the firms, competition for a secure place in the world market has become much more acute. Managers' attitudes are adjusting as they seek new sources of competitiveness, either by internal development or in partnership. Their operations reflect, as in a mirror, the growing interdependence of national economies perceived by their opposite numbers in national governments. Few can afford to ignore the developing world, either as a market, or a source of supply, or indeed as a source of new competition. As buyers, developing countries as a whole account for a quarter of world imports, the majority of which comes from developed countries. For many multinationals, such as Komatsu in earth-moving machinery and GEC in electrical machinery, trade with developing countries can spell the difference between profit and loss. As suppliers, they provide nearly 30 per cent of world exports, 60 per cent of which is in manufactures, the more capital-intensive parts of which go to developed countries.[5] And as third world multinationals gain in stature, they add further pressure to change the rules of the competitive game. Most multinational firms recognise the importance of these issues, even those who remain unwilling to invest in developing countries.

Similarly, most developing countries feel the pressure to re-examine how foreign firms with their command of finance and technology and their access to rich markets can help offset the dire consequences of the failure of sovereign borrowing to provide a reliable and secure engine of growth. As we show in chapter 4, many have relaxed their previously stringent criteria for screening potential investors and added more generous incentives. By their actions, they are showing how far opinion has moved from the early 1980s, when one study could report that three-quarters of diplomats thought that all multinationals employed corrupt practices and policies detrimental to development.[6] They are also rejecting the conclusions of many quite recent academic studies, one of which asserted that 'the multinationals have undermined local economic and social autonomy' (Dixon *et al.*, 1986: p. 16). It is not the firms *per se* that have constrained autonomy; it is the world system in which the firms are but one set of players and are increasingly recognised as such.

There is a new hope that firms can provide poor countries with both the stimulus and some of the resources they need for economic development without unduly undermining national self-esteem. The dynamism of global structural change offers states new options and makes the future seem altogether brighter. Yesterday's straitjacket of fixed opinion can be shrugged off. Yet there are risks and uncertainties. As the next chapter shows, structural change both helps and hinders aspirations for development. Moreover, beneath the veneer of growing pragmatism and collaboration, there often lurk fears that the old possibilities for exploitation by foreign interests remain as strong as ever. Many government officials to whom we talked expressed the hope that their countries would eventually be able to break away from the current reliance on foreigners to promote their exports. The necessity they feel to dash for export-led growth can call for bedfellows that prove only temporary.

Optimism about a new rapprochement can thus be overstated: a latent hostility may remain. Besides, it is perhaps ironic that while the official climate for inward investment has improved, the investment flows into most developing countries have declined. In part, this seems to reflect the fact that most of the policy changes address the problems of the 1970s and ignore the opportunities of the 1980s. In the service sectors, where much of the new dynamism lies, old attitudes and restrictions seem as firmly implanted as ever. This is an issue that we address at the end of the book when we have shown the evidence for both optimism and pessimism about the possibilities.

Where does government policy fit in?

Liberalisation of policy combined with falling costs of cross-border transactions has rekindled old fears in many quarters that the multinationals are becoming more mobile and more stateless world citizens, divorced from the competition-distorting effects of national policy. This depiction of the multinational has been promoted by such observers as Ohmae (1990), who point to some of the outcomes described in chapter 3. In our opinion, however, Ohmae's view is perhaps no more than a portent of the long-run future: closer inspection of the evidence shows that only a few firms can hope to operate in a 'borderless' world. Governments, both host and home, continue to play a crucial, and perhaps paradoxically, an increasing role.

Porter (1990), in particular, has advanced strong arguments to show that firms draw their vitality from the conditioning forces in their home markets. He asks why some nations are more prosperous than others and why some national sectors flourish while others stagnate. He focuses attention on four factors. One is the basic structure of national factor costs, including the supply of skilled workers and an efficient infrastructure. For this he draws on earlier work in trade economics to question many of the basic assumptions of comparative advantage (Ohlin, 1933). Another is the structure of demand conditions, affected by national macro-economic policy and in turn affecting the composition of trade in ways shown earlier by such scholars as Linder (1961). Equally important are the effects of competition (echoing Schumpeter, 1942) and the impact of related and supporting industries (familiar to students of Marshall in terms of 'externalities'). His central thesis is that competitiveness is born of fierce local rivalry, an active anti-trust policy and avoidance of protectionism. 'Competing domestic rivals will keep each other honest in obtaining government support. Companies are less likely to get hooked on the narcotic of government contracts or creeping industry protectionism'. Despite the title of Porter's work, *The Competitive Advantage of Nations*, the main lines of argument apply at the level of the firm. No country can be good at everything, for only some firms and some sectors thrive within the framework established by economic management and the workings of a national culture.

Porter's analysis provides only a point of departure for many of the lines of argument advanced here. Though the basic analysis is contained in the older ESP paradigm (Koopmans and Montias, 1971), his analysis takes life from the detailed analyses of contemporary conditions. None the less, he omits detailed consideration of entrepreneurship and investment. More seriously for our purposes, he almost wholly ignores the changes in the world system outside the countries and fails to recognise the composition of government as groups of parties with different interests. He also fails to examine the interaction between the international competitiveness of local firms and government policy; a crucial issue in many developing countries (Aggarwal and Agmon, 1990). Yet his analysis can be adapted to host, developing countries. For this one needs to add more explicit treatment of government policies that balance economic with social conditions. Small, poor countries cannot afford the luxury of letting market forces determine outcomes. Besides, even in the USA, it is recognised that government action can alter the balance of factor cost advantage through intervention (Zysman and Tyson, 1983). Adding explicitly the role of government policy and the shifts of global competition to Porter's four factors produces the diagram of related influences shown in figure 1.1 below.

Rather like Toynbee's concept of challenge and response, the diagram suggests the possibility of an explanatory framework: it is not the basis of a deterministic theory. Its meaning emerges from the interactions among the variables, more in terms of what is contradicted than definitively established. By suggesting new questions, it can help articulate, but not resolve, the long-standing debate about the lines of causality between policies to create growth and those aimed at structural reform.

Causality can be seen to run in both directions. Governments may either take the initiative to influence the composition of output or trade, or feel forced to respond to external changes they regard as undesirable.[7] There is therefore great uncertainty about the appropriateness of conventional prescriptions for development policy. Besides, the ability of forecasters and policy makers to pick 'winners' is low. Who would have forecasted in the 1950s that South Korea would maintain a real growth rate of 5.7 per cent for thirty years? More likely, economists would have looked at the low savings rate and considered South Korea a likely loser (Crook, 1989). Malaysia would probably have been written off on the same basis, especially given its security problems at that time.

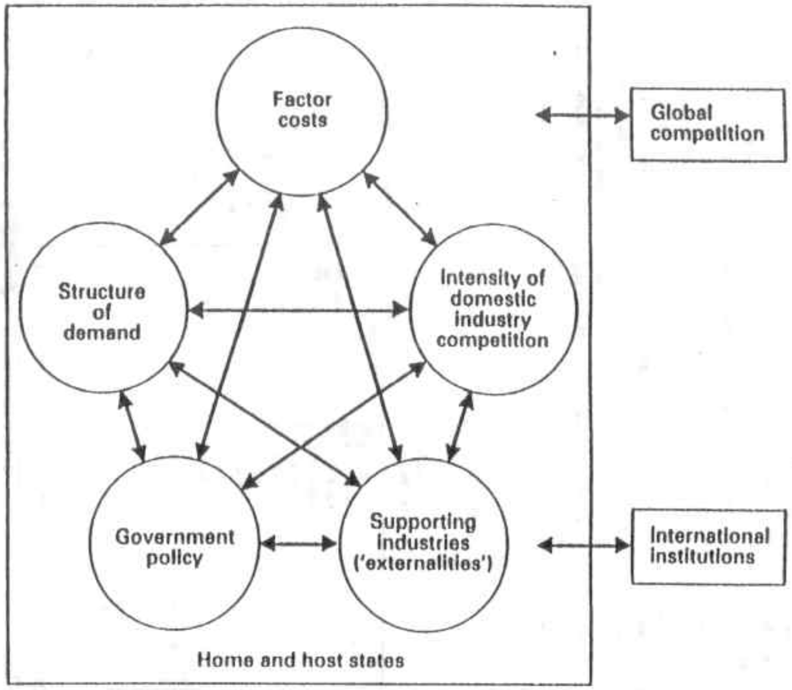
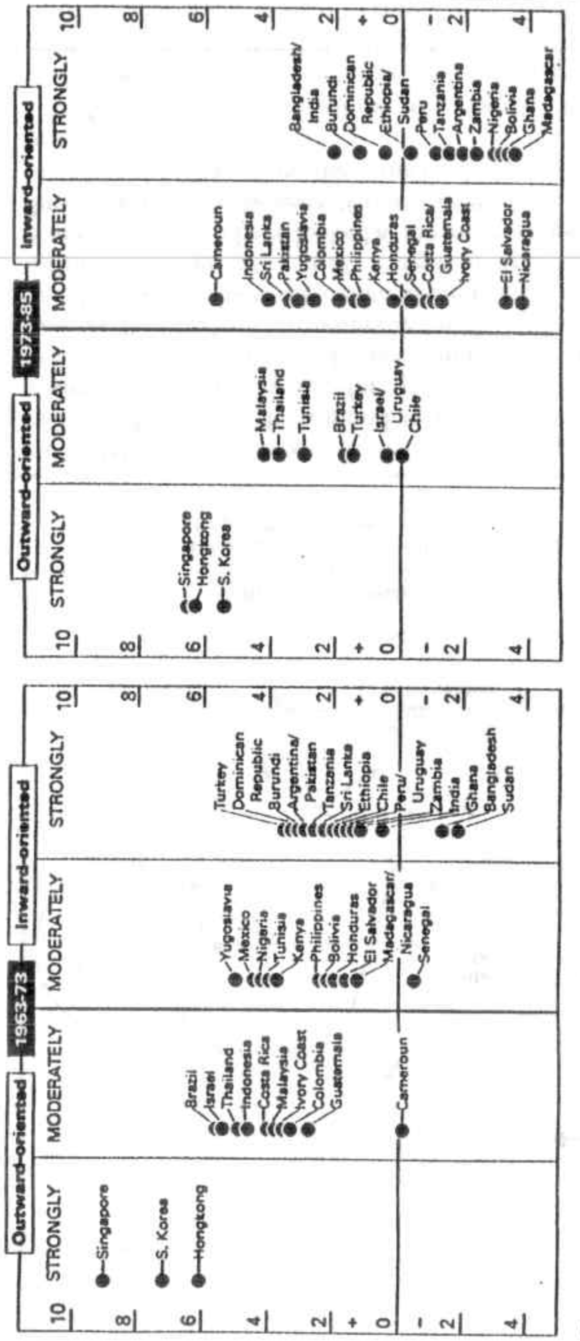

Factor costs
Global competition
Structure of demand
Intensity of domestic industry competition
Government policy
Supporting industries ('externalities')
International institutions
Home and host states

Figure 1.1 Primary factors influencing a nation's competitiveness. Adapted from Porter (1990).

By focusing on the role of the firm, figure 1.1 can also be used to challenge the theory of comparative advantage that has formed the basis of so much official thinking. Many economists have begun to develop new models that accord more closely with actual trade behaviour.[8] As Krugman (1987: p. 131) has said, 'these new models call into doubt the extent to which actual trade can be explained by comparative advantage: they also open the possibility that government intervention in trade . . . may under certain circumstances be in the national interest after all'. For example, in capital-intensive industries such as chemicals, the decision to build a new export-oriented refinery can disturb existing trade patterns. The location of such lumpy investments is chosen as much by the investor's calculation of advantage relative to all its existing refineries as by the national factor costs and demand function in the host country. Though we do not set out to build an omnibus theory that can embrace such calculation, we recognise that trade and national advantage must be seen in the context of the bargain struck between the investor and the host nation. Countries now compete for scarce managerial resources more by providing a favourable investment 'climate' than simply relying on factor cost advantages.

A further issue in any analysis of policy choice is the *power* of the state to implement its choices and overcome social, religious and political obstacles to change. How Brazil and South Korea responded to the problem of spiralling debt and reduced inflows of foreign capital illustrates the impact of domestic circumstance on policy and thus on the investment climate. In 1982, the Brazilian government, afraid of the social effects of the decline in capital inflows, cut investment even more severely than they cut consumption. By contrast, the Korean government, more confident of its ability to contain and suppress social unrest, acted to increase domestic saving and stepped up domestic investment to take the place of foreign loans, even though this meant a cut in real wages and consumption. Growth rates in Korea stayed high: they declined in Brazil. Brazil's choices in the 1980s reflected in part the relative weakness of its government in the aftermath of the transition to civilian rule. Weak states usually resort to approaches that can be regarded as 'making policy by default'.

Related to these choices of internal policy are those affecting attitudes towards 'outward-looking' policies. Despite the claim of bodies such as the IMF, some countries hold the view that an 'outward-looking' policy framework does not always promote growth and that inward-looking approaches can be superior. Lall (1987: p. 28) asked, 'can Tanzania become Korea only by adopting an outward-looking trade policy?'. Clearly not, when the necessary technical abilities and entrepreneurial drive is absent. The spread of growth rates under each policy regime shown in figure 1.2 below supports such agnosticism and illustrates the difficulty of drawing general conclusions. As the 1989 United Nations Conference on Trade and Development (UNCTAD) trade and development report stated,



Strongly outward-oriented
Trade controls are either nonexistent or very low in the sense that any disincentives to export resulting from import barriers are more or less counterbalanced by export incentives. There is little or no use of direct controls and licensing arrangements, and the effective exchange rates for imports and exports are roughly equal.

Moderately outward-oriented
Incentives favour production for domestic rather than export markets. But the average rate of effective protection for the home markets is relatively low and the range of effective protection rates relatively narrow. The use of direct controls and licensing arrangements is limited. The effective exchange rate is higher for imports, but only slightly.

Moderately inward-oriented
Incentives clearly favour production for the domestic market. The average rate of effective protection for home markets is fairly high and the range of effective protection rates relatively wide. Direct import controls are extensive. The exchange rate is somewhat overvalued.

Strongly inward-oriented
Incentives strongly favour production for the domestic market. The average rate of effective protection for home markets is high and the range of effective protection rates wide. Direct controls and licensing disincentives to the traditional export sector are pervasive, positive incentives to nontraditional exports are few or nonexisting, and the exchange rate is substantially overvalued.

'export reforms are no guarantee of economic growth'. Besides, some of the external choices clash uncomfortably with domestic priorities. Developing effective external policy to gain internationally calculated *efficiency* to support an export drive can conflict with domestic priorities for *equity* or racial justice.

Resolving such clashes cannot be solved by rational economics alone; they must be regarded as political choices. An interview with India's prime minister in 1990 provided an illuminating vignette of the dilemma. Mr Singh had publicly acknowledged his awareness that the events in Eastern Europe and the USSR were shaking India too, and that India risked being left further behind if it did not accept the new international challenge. When asked why he had not liberalised and deregulated faster, he replied, 'We are politicians. We know our place. The people have to be persuaded first: they will understand what is confronting them when it is explained . . . but it will take time.'[9]

India's long-run performance has been much poorer than South Korea's. In 1950, both were miles behind the industrialised countries and it seemed then almost impossible for either to attain high standards of living. Today India has a per capita income of about $250 (in 1980 $) and South Korea's is roughly $2,900. Similar differences show up in Korea's far greater levels of social gains, such as greater life expectancy, education and change in the structure of the domestic economy. What caused the difference?

Though it might be tempting to attribute South Korea's greater success to its outward orientation and India's problems to its inward-looking priorities, the many differences in social and economic conditions make it impossible to provide a single answer. Besides, aggregate statistics run the risk of provoking sweeping generalisations; they obscure sectors of relative growth or relative decline. None the less, there are some important similarities and contrasts that seem to bear on the issue. For instance, both governments followed policies of strong intervention in their markets – an outward-looking policy does not necessarily mean *laissez faire*. The crucial difference is that India constrained competition, whereas South Korea actively promoted both domestic rivalry and the international development of local firms.

There is a balance to be struck. As Lawrence (1987: p. 102) said, 'an industry needs to experience rigorous competition if it is to be economically strong. Either too little or too much competitive pressure can lead an industry to a predictably weak economic performance characterized by its becoming inefficient and/or non-innovative.' This sense of required balance is a repetitive theme in the chapters that follow, for countries need to steer carefully between the scylla of many-headed foreign competitors (as in Brazil) and the charybdis of government protection (as in Kenya).

We believe that sustained investments in building an educated society whose *mores* are those of competition (as in Japan) provide the best means of steering a middle course and of achieving a quality of growth that is sustainable even in the face of external shocks. Sustainability depends largely on underlying skills, on social equity, and on the avoidance of the careless exploitation of natural resources and environmental damage that will harm later generations. The Ivory Coast illustrates how temporary success can be. For a time, it was the darling of the development agencies as its outward-looking policies generated growth. So strong was the economy in 1979 that, for example, one large multinational made 10 per cent of its world-wide profits there, largely because of its ability to supply neighbouring Nigeria with a wide variety of standard products that could not satisfactorily be produced under the then chaotic Nigerian conditions. Those exports have now dried up as Nigeria has put its own house more in order. Moreover, the Ivory Coast placed much reliance on the same commodities as its neighbours, thus entering a zero-sum game that worsened its plight when prices fell. And its exchange rate, pegged to the French franc, turned from an advantage to a constraint in the ensuing phase of enforced adjustment. From looking outwards, the Ivory Coast has now turned inwards. Robustness of policy combined with an ability to adapt systematically to changing external circumstance are qualities just as important for states as they are for firms.[10]

THE GROWING REACH OF THE MULTINATIONALS

International business is an ancient feature of the world economy. The Phoenicians opened up new trade routes and stimulated new sources of production. Indeed, the travelling merchant has been one of history's unsung heroes in providing the initial contacts among entire cultures. It is only in the last century that the multinational corporation has emerged in forms we would recognise today. Led by pioneers such as the American firm, Colt, which established a revolver factory in Britain in 1852, these firms began to change the old order of trade and finance as the primary weapons of international business. They added international production of goods and services and began to manage their overseas assets as part of an integrated unit rather than as passive investments in a portfolio.

More recently, there has been an explosion of FDI (foreign direct investment). While world trade volumes grew at a compound annual rate of 5 per cent between 1983 and 1988, FDI grew at 20 per cent annually in real terms. Why this occurred has much to do with the structural changes described in chapter 2. The consequence has been to tie more closely together the determinants of trade and investment flows. A milestone was reached in the mid-1980s, when the volume of international production for the first time exceeded the volume of international trade. The increased FDI flows have, as Julius (1990: p. 36) claimed, 'reached the threshold where they create a qualitatively different set of linkages among advanced countries . . . [A]s a means of international economic integration, FDI is in its take-off phase, perhaps in a position comparable to world trade at the end of the 1940s.'

The change is significant as it diminishes the power of states to control economic events. States retain considerable *negative* power to disrupt, manage or distort trade by controlling entry to the territory in which the *national* market functions. They cannot so easily control production which is aimed at a *world* market and which does not necessarily take place within their frontiers. And even when most of the supply is under control, the market may not be, as OPEC discovered to its cost in the 1980s. In other words, states' *positive* power to harness internal resources is decidedly constrained when they try to influence where and how international production takes place. They find they cannot direct; they can only bargain. And the costs to themselves can be very much higher than those of trade protection. For example, when the Brazilian government introduced its Informatics Laws to prevent the international computer enterprises from producing in Brazil, it imposed a very high cost on those local enterprises needing to use the latest and best computers in order to keep up with *their* competitors. The outcome of bargaining may be to the state's advantage, or to the multinationals'. It is still a bargain.

How powerful the multinationals have become is a notoriously difficult question to answer. It is not solely a question of the statistics on FDI, the asset base of the multinationals.[11] These understate the growing reach of the multinationals by ignoring other ways these firms affect performance. Yet much official attention concentrates on these figures. In the hey-day of sovereign lending to developing countries, when FDI accounted for less than 10 percent of the resource transfers in some years during the 1970s, many thought that the multinationals were of little consequence for development. Figure 1.3 shows the trends as measured by the IMF. An even smaller role might be accorded to the multinationals as a source of finance if one deducts the remission of profits from the FDI flows. In most years during the 1980s, the IMF data show that the 'real' transfer has seldom exceeded $3 billion.[12] Perhaps this was why the authors of the Brandt Reports (1980, 1983, 1985) ignored almost totally the phenomenon we address in this book.

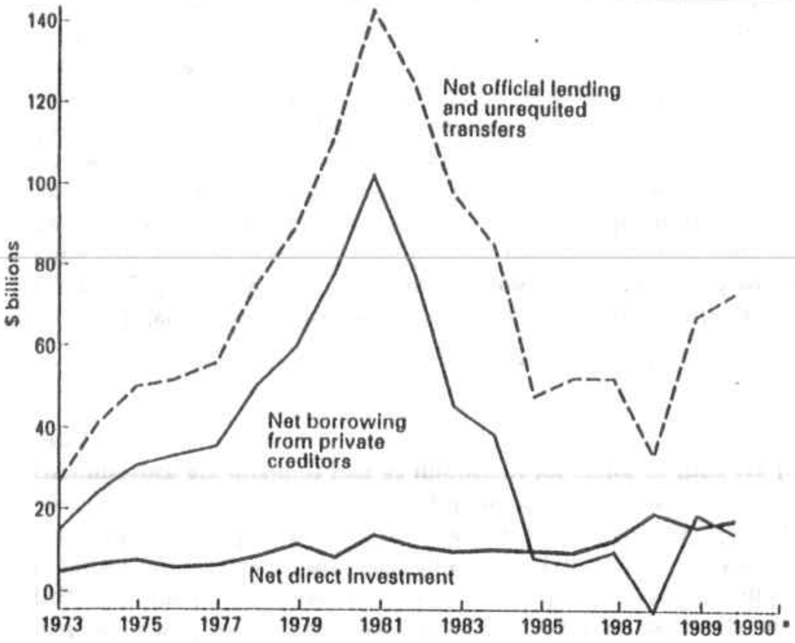


Figure 1.3 Financial transfers to indebted developing countries, 1973–1990 (excluding reserve-related liabilities). Adapted from *World Economic Outlook* (Washington D.C.: International Monetary Fund, various issues).
* Forecast

There are three basic reasons why such statistics misrepresent the role of the multinationals in development. The first is that the multinationals' share of and influence over world production and trade is far greater than their share of the capital transfers. Directly or indirectly, they have been estimated to control over one quarter of the world's economic activity *outside* their home countries;[13] over half the world trade in manufactured goods and even more of the growing trade in services; 80 per cent of the world's land cultivated for export crops, and the lion's share of the world's technological innovations. The accumulated stock of FDI was estimated to have exceeded $1 trillion by the end of 1988. Table 1.1 gives some detail on the growth of

FDI, where it originates and where it is invested. The picture is one where the concentration of activity is increasingly within the developed world, leaving all but a few developing countries outside the reach of the new dynamism. The early post-war dominance of the USA as the source of FDI has been reduced as Europeans and the Japanese have invested in the USA. Indeed, if one excludes US-sourced FDI, the USA alone has attracted 60 per cent of all FDI flows during the 1980s. The declining US share of FDI is leading to a more diffused structure of economic power and a more fiercely contested international market. Though these estimates should not be taken too literally, they indicate the orders of magnitude of the multinationals' activities.

Impressive as $1 trillion sounds, the real asset base is even higher if one takes into account the effects of debt and the indirect effects of the non-equity and contractual forms of association, the so-called New Forms of Investment (NFI) that have been growing rapidly in recent years. For example, Reuber (1973) found in the years before the bank-lending boom that 54 per cent of the capital raised for the sixty-eight projects he examined was in the form of debt and that 40 per cent of the equity was raised in the developing countries. Thus, the total resources affected by the multinationals far exceeded the total suggested by their equity base. Though some observers hold the view that NFI opens the door to more equitable arrangements (Oman, 1984, 1989), recent analysis suggests that, for many classes of activity, these arrangements are inherently inefficient (Hennart, 1989). NFI may be no more than a second-best solution to distortions in international markets, capable of alleviating some problems but worsening others. Much depends on the individual firm, whether it is defending or attacking an established market position. The issues are discussed in chapter 5, for they have a large bearing on attitudes that form the basis for policy choice.

An alternative measure of the multinationals' significance is in terms of their employment. They are estimated to account for only 10 per cent of employment in developed countries. In developing countries, they employ perhaps twenty-one million men and women directly. When indirect employment provided through subcontractors and suppliers is included, the International Labour Office (ILO, 1984) estimates this still only doubles the employment base, to perhaps 2 per cent.[14] These estimates overlook the influence exerted on host economies through multiplier and similarly indirect mechanisms.

The second reason for underestimating the multinationals' reach is their growing role as agents of integration with the world system. A large and growing share of merchandise and financial trade is managed within the firms' networks, either as part of their operations or ancillary to the rapidly expanding service activities both within the manufacturers and among service specialists. Data from many nations show that over 40 per cent of their external trade in manufactures is managed as transactions among firms' affiliates and that the figure could be increased by adding in the trade among multinationals.[15] For example, over 80 per cent of the technology payments received by the United States are royalties paid by affiliates to their US parent companies.

Table 1.1. *Stock of Foreign Direct Investment, 1960–1988 ($ Billions)*

	1960	1971	1980	1987	1988
A. By source country					
United States	33	83	215	314	334
United Kingdom	10	24	74	179	184
Netherlands	7	14	40	79	70
W. Germany	0.8	7	38	100	97
Japan	0.5	4	37	77	111
Other developed countries	15	36	93	179	199
Developing countries	0.7	4	14	40	45
Total (estimated)	67	172	512	968	1,040
B. By host country					
United States	8	14	68	272	328
Canada	13	28	46	100	N/A
United Kingdom	5	13	45	95	110
Other developed countries	11	60	165	N/A	N/A
Developing countries	18	51	117	N/A	N/A
Total	55	166	441	N/A	N/A

Note: The data for source and host countries differ due to different national reporting procedures; annual estimates vary according to currency translations into US $.

Source: These data are drawn from many sources. The principal ones are national official sources, plus Stopford and Dunning (1983) (for 1960-1980 data); JETRO, White Paper on Overseas Investment (annual series); UNCTC (1988a); and authors' estimates.

Table 1.2. Concentration of flows of FDI to developing countries (% share of worldwide inflows)

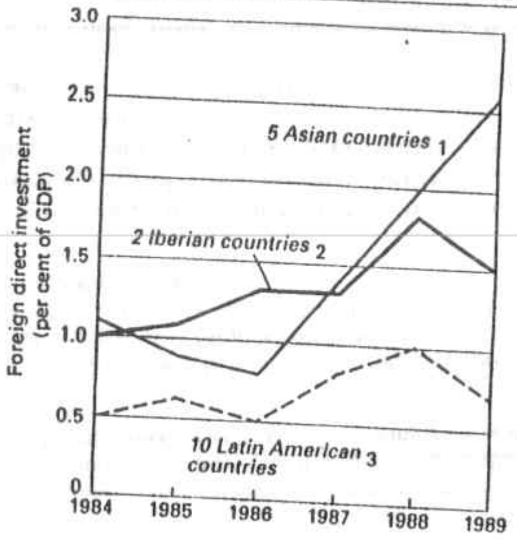
	1981–3	1984–7
All developing countries	27.5	21.2
Latin America and Caribbean	13	8
(of which, Argentina, Bermuda, Brazil, Colombia & Mexico)	(82)	(93)
South and East Asia	11	9
(of which, China, Hong Kong, Malaysia, Singapore)	(78)	(72)
Africa	1	1
(of which, 8 oil-exporting countries)	(70)	(83)

Source: UNCTC, (1989), 'FDI Flows in the Mid-1980's', UNCTC Reporter, No. 27, Spring.

These 'hidden' consequences of the spread of FDI are part of the general trends that limit the ability of governments to control events.

The third reason for not taking the FDI figures too literally is that, like all aggregates, they hide more than they reveal. As table 1.2 shows, a few countries attract the bulk of the investment in developing countries. Of the total flow during the 1980s, nearly 90 per cent went to just twenty countries, compared to 70 per cent in the early 1970s (UNCTC, 1988a; IMF, 1985). That concentration is increasing as multinationals seek the best prospects for growth and focus their interest on those countries with abundant natural resources, large internal markets, plentiful skilled but relatively cheap labour, and favourable regulatory regimes. This trend reflects the fact that the multinationals tend to be influenced by and in turn influence the ability of countries to grow.

The multinationals' impact on those countries receiving the greatest investment attention has been considerable relative to GDP, as figure 1.4 indicates for a few countries. Yet these aggregates hide major sectoral differences. These are due as much to regulatory variations as to availability. In Brazil, for example, 62 per cent has been in manufacturing, whereas in Chile, services together with mining and oil have taken the lion's share. Moreover, the accumulating stock of investments have produced quite different impacts on the proportion of sectors controlled from abroad. Thus, while over 60 per cent of Singapore's manufacturing is foreign-controlled, in India the equivalent figure is well below 10 per cent.[16] In short, generalisations obscure the dynamics of adjustment. Why this should have been so and how these firms might affect the course of development in the future is the focus of all that follows.



1 Indonesia, Malaysia, Phillipines,Singapore, and Thailand.

2 Portugal and Spain

3 Argentina, Brazil, Chile, Columbia, Costa Rica, Equador, Mexico, Peru, Uruguay, and Venezuela

Figure 1.4 South-East Asia lures investors. Adapted from 'Fostering Foreign Direct Investment in Latin America', mimeo, Institute of International Finance, Washington D.C..

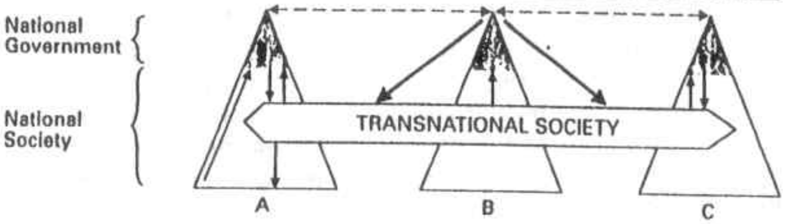
TRIANGULAR DIPLOMACY

Cross-border competition is intensifying for both firms and states. For firms, decisions in one country are now more frequently influenced by choices made elsewhere. For example, Firestone, the US tyre producer, renegotiated its position in Kenya in part because of the competitive battles it was fighting in the industrialised markets. As we show below, the effect of global competition on Firestone's cash flows acted powerfully to change managers' opinions about previous policies of ownership and control. States are also competing among themselves to woo new investors. Malaysia, for example, has stated publicly that its recently enhanced incentive packages are calculated with a clear knowledge of what other governments in the region are offering.

The ideas that interdependence necessarily involves closer interaction among governments and firms is not new; it was suggested nearly twenty years ago by a German scholar, Karl Kaiser, when the debate first opened on the meaning and significance of 'interdependence' (Kaiser, 1971). Implicitly criticising much of the then-current discussion, Kaiser pointed to the asymmetry of government power in transnational politics. He drew on the work of Aron who had introduced the concept of 'transnational society' to international relations to indicate that states' options are affected by developments in the flow of ideas and beliefs across borders and by non-national organisations (Aron, 1966; p. 105).[17] To this he added Perroux's ideas on dominant economies (1950) and drew a diagram (shown in slightly modified form in figure 1.5 below) to illustrate how all these ideas could apply to the power and influence of foreign companies in less developed countries. His arrows indicate attempts by one group to influence others. Some point downward, implying direction, coercion, regulation, rather than dialogue. But they also point upward, implying some reverse influence of business on government within countries.

Kaiser's diagram shows only a two-way relationship to indicate that some states have more influence than others over the conduct of international organisations and firms. What he omits is the nature of the interactions within his 'transnational society' and how these affect the options and power of states. More recent developments have shown that these interactions have developed strongly both to add a third dimension to the phenomenon and to suggest further limits to the power of the state.

These interactions are indicated by the difficulty states increasingly have in controlling population movements: witness the illegal immigrants in the USA and southern Europe. Still less can they control the flow of ideas. Much of the recent change in the Eastern Europe and the USSR has been sparked by individuals' heightened awareness of their relative material backwardness, let alone political constriction. Rising consumer expectations in developing countries add pressure on governments barely able to cope with basic needs. Labour unions have also exported their ideas. Though their attempts to organise labour internationally, as in the International Metalworkers Federation, have not amounted to much, their ideas are pervasive.



A, B, C stand for nation states: see text for definition of 'transnational society' and the arrows.

Figure 1.5 .Transnational politics. Adapted from Kaiser (1971).

Of particular note has been the emergence of a privileged transnational business civilisation, or what Sunkel (1973) labelled the 'transnational kernel'. Its US core is in New York – not Washington – and in Chicago and Los Angeles, where the social and political elites most wholeheartedly accept the values, the *mores*, the customs and taboos of the business civilisation. So do their counterparts in London, Tokyo, Sao Paulo, Sydney and Taipei. Their values are both economic (efficiency, speed and responsiveness to changes in demand) and social (openness to competition, opportunity for social advancement regardless of race and parentage, and soon, perhaps, of sex). Particularly on the latter issue, practice, as always, falls short of ideology. Women are not yet given equal opportunity; nor are blacks, as a headcount of business class passengers on any airline will show. But the significance of each of these values is that the core is in the lead in bringing about social and economic change. American women have more opportunities than Japanese women; American blacks more opportunity than Japanese of Korean origin. There seems little doubt that where the core has led, the business civilisation will follow. Far from acting to preserve dependency in small 'peripheral' states, even radical analysts now seem to be accepting that this privileged class can play a crucial and positive role in economic transformation (Weiss, 1988: p. 159).

These three sets of forces are transforming the old game of Diplomacy. No longer can national boundaries define the rules, for the game is now one where negotiation and action is carried out on a triangular basis. The traditional players in the embassies and foreign ministries are still in business, but they have been joined by members of other government ministries and by the executives of firms, both local and multinational All are now involved in both bilateral and multilateral negotiation. Figure 1.6 shows the three dimensions of the game in a simple form. Later on, in chapter 4, we complicate the

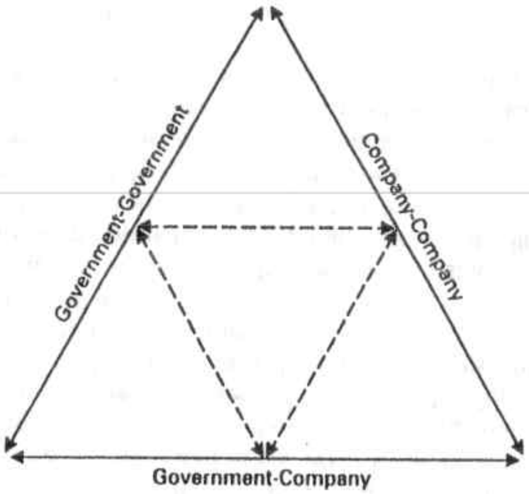


Figure 1.6 Triads of relationships.

diagram to distinguish the nature of the relationships affecting local and foreign firms.

Changes in the international political economy create new sources of asymmetry in the relative importance of the three sides of the triangle. The growth of global competition can be seen as moving the world towards a position where events are coiditioned more by an emerging managerial technocracy than by traditional notions of state power. In this new technocracy, firms feature prominently but are only one component of a wider network that links them to the educational and skills infrastructure and the financial system. Competition is increasingly among different production and institutional systems and contrasting social organisations. Further challenges to the role of the nation state are provided by the Europe of 1992 and North American debates about a free-trade area to run from Canada to Mexico. Supra-national bodies could emerge to offset the growing power of the world market. We are not attempting to forecast from among these and other possible 'meta' structures; we merely indicate that new triangular relationships will continue to evolve.

In the explicit bargaining between host governments and foreign enterprises, sometimes one and sometimes the other has the whip hand. Frequently, as in inter-state diplomacy, elements of both conflict and cooperation exist simultaneously in the bargaining process.[18] The asymmetry of the process makes the choice of partner critical. For states, a sickly corporate partner may be just as damaging as an alliance with a weak or vacillating state. The other side of the coin is the firm's choice of national site. Before the 1980s, corporations looking for compatible hosts shopped for tax breaks and docile, low-paid labour. More recent experience has demonstrated that neither could compensate for administrative incompetence or for official intervention, either by internal price control or tariffs or other barriers to efficient operations.

Recent scholarly interest in economic organisation has spawned a controversial view of the modern corporation as a 'nexus of contracts' (Aoki et al., 1990). We have gone further, regarding state–firm bargaining as the nexus of international treaties affecting the balance of power; the lasting effectiveness of the deal is determined by the success or failure of bargaining on all three sides of the triangle. When we examined the record of many investors in our three countries, we found that we could not understand much of what had been decided and done without explicitly taking into account shifts elsewhere. If one looks at only one side of the triangle, one misses important factors.

Three examples make the point. Some years ago, Britain infuriated the Indonesian government by reducing its (already small) import quota for T-shirts as part of one of the endless rounds of renegotiating the Multi-Fibre Arrangement. Indonesia sought ways of retaliation and found that one British firm, Davy International, was constructing a large chemical plant in their country. They embargoed the project as a way of voicing their displeasure and put at risk a contract worth many times the value of the trade in T-shirts. Davy might have foreseen the risk had they been monitoring the state–state negotiations, but who in that industry then would have considered the cost of such monitoring worthwhile? Malaysia also illustrates how unexpected risk can be created on another side of the diplomatic 'triangle'. Malaysian anger over British 'colonial' attitudes and especially over the decision to raise the fees for foreign students in Britain led to Dr Mahatir's 'buy-British-last' policy. British exporters were affected much more than the investors, but both were hurt by the consequences of a state–state dispute that had little, at least initially, to do with economic relations. A third example concerns Kenya. When the British Foreign Secretary announced that British aid would be reduced to countries abusing human rights, the British High Commission in Nairobi had to lobby strenuously to get the UK government to back off, so as to protect British investors' interests.

A MATRIX OF AGENDAS

To begin to capture the sense of this complexity, we tried to model our initial propositions, but found that we could not cope systematically and simultaneously with all three sides of the triangular relationships. We therefore concentrated on the direct relations between states and firms, leaving considerations of state–state and firm–firm bargaining as modifying factors in specific circumstances. The result was a simple matrix that provides a snapshot of the range and scope of the interests of both sets of players at a given time. Figure 1.7 shows nine possible combinations of circumstance. The concentration is on trade policy as a proxy variable for other critical issues affecting negotiations. Though this is a simplification of reality, we chose to focus on trade, as export creation is now one of the most urgent needs of developing countries. The matrix is not, however, designed to define all possible roles for multinationals in development, but to identify some of the major arenas in which bargaining takes place. We complicate the model in later chapters, most particularly in terms of the dynamics of change on either axis and in terms of issues that reflect the interactions with the other two sides of the triangular diplomacy, but for now the simple version can serve to introduce the ideas.

Governments are placed on the horizontal axis in figure 1.9 in terms of their trade policy 'Intent'. They are considered to have three primary options. One is to regard inward foreign investment as a means of replacing imports and to provide encouragement for that by offering tariff protection or other barriers to competition. This was the traditional view, dominant in the 1950s and 1960s. Developing countries typically used import-replacement as an initial step to create the industrial infrastructure needed for later exporting (Chenery et al., 1986; Page, 1986; Stewart, 1982). Yet, some economists have argued that to equate industrialisation with import-substitution has been mistaken, because of the high costs of the transformation needed later on (Krueger, 1983; Pazos, 1985). This is an issue to which we return in chapters 4 and 5, for both states and firms find that eliminating the bad habits engendered by operating behind protective barriers is far from easy, and sometimes impossible. In many cases, national firms cannot change for lack of resources and multinationals are resistant to change for fear of upsetting arrangements elsewhere. None the less, many states would consider they had no other initial option.

More recently, many developing countries have actively sought to add export-creating investments without going through the stage of import-substitution. They have done so in a great variety of ways, but all have been spurred by the need to earn hard currency to finance domestic investments and service payments on prior loans. We draw a critical distinction between policies designed to replace imports and those for export-creation, for they have much to do with the questions of efficiency, appropriate technology, managerial competence and the likely effectiveness of alternative forms of incentive.

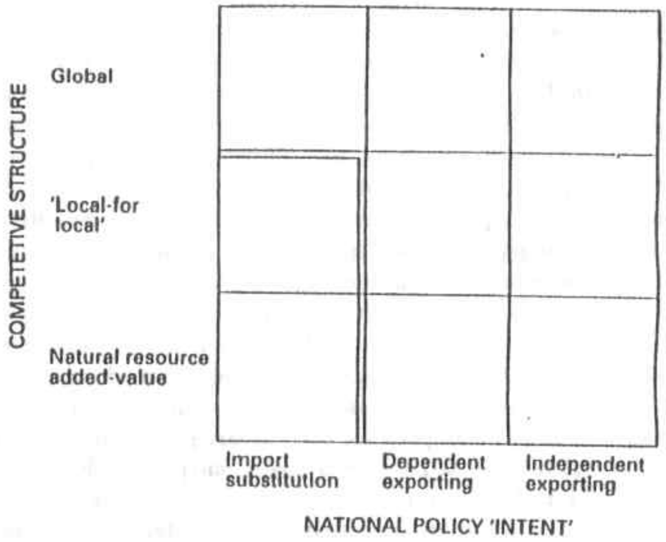


Figure 1.7 Nine possible bargaining sites.

We also draw a distinction between two basic approaches to creating exports. A country can recognise its reliance on foreigners, either by ownership or by contract. Multinationals possess power by virtue of prior investments in distribution systems to serve the world market. Equally, states can benefit from the power of distribution groups in industrialised countries who offer long-term or seasonal contracts for supply of goods at pre-determined specifications. Sears Roebuck in the USA, for example, has for long supported local entrepreneurs in South-East Asia by providing engineering drawings and even assistance in factory layout for a wide variety of products. Some contract farming of cash crops has similar features. In addition, manufacturing companies like Amstrad in consumer electronics and Courtaulds in textiles now have wholly locally-owned factories producing exclusively for them. Such exports are labelled as *dependent*, for they will only continue to exist for so long as the foreign entity considers them economically viable.

By contrast, some countries like South Korea have actively sought to promote exports from indigenous producers. They have valued *independent* exporting for the freedom it gives them to determine their own position in world markets. Malaysia sought some of these benefits by buying the parent companies of politically important subsidiaries, such as Sime Darby, which now promotes itself as 'South-East Asia's largest multinational'. Whereas *dependent* exporting allows countries to avoid the high entry costs to international selling and distribution, policies designed for *independent* exporting explicitly recognise these costs and the associated risks that local firms must bear if they are to reach and sustain the required levels of efficiency and performance.

On the vertical axis in figure 1.7, firms are considered to operate in one of three basic structures of competition. Most manufacturing and service industries can be regarded as having either of two configurations of international activity. One, labelled as 'local-for-local', refers to those industries where the primary locus of activity is close to each market, with relatively little interaction with other markets. Accounting services and toiletries are examples, even though the transfer of skills and the firm's international reputation or brand are important. 'Local-for-local' production was the predominant structure during the early expansion of FDI up to the 1960s. Many industries have subsequently shifted into a 'global' configuration, where the interaction among markets is much greater. These interactions involve trade in components as well as finished goods together with the use of cash-flows and managerial systems as competitive weapons for competing across borders. Electronics, automobiles and specialised machinery are examples. There is a third category of competition, the local added-value processing of natural resources. We exclude the production of basic commodities and concentrate on later stages of the processing, for which there are choices as to the location of the plants. The power of the local country to affect decisions about such added-value processing is considerable and generally greater than for other types of manufacturing or services.[19]

The bargaining agenda in each of the nine 'cells' of the matrix is subject to different balances of consideration. Table 1.3 provides a crude identification of some of the important variables affecting which side holds power in the negotiation. The reality is, to be sure, more complex. One problem is that relative power changes *over time*. Governments may hold power to control and regulate the outcomes when a firm first enters the country, especially if there are many competitors who can be played off one against the other, but may lose influence later on as operations become established. Conversely, in risky projects such as exploring for oil or other minerals, the firm may hold the initial advantage but lose it once the project has been successfully completed. This, latter, case is what Vernon (1977) has labelled the 'obsolescing bargain'. As we explore in later chapters, such obsolescence does not occur so markedly in other sectors.

Table 1.3. *Factors influencing relative bargaining advantage*

	Advantage held by	Likely regulation
Factor		
Dependence on local resources	Government	High
Dependence on local markets	Government	High
Political 'salience'	Government	High
Industry structure		
Many competitors	Government (at entry)	Variable by product
Global 'linkage'	Firm	Low
Business dependent on:		
Proprietary knowledge	Firm	Low
Highly complex process	Firm	Low
Labour-intensity	Government	Low–high
Capital-intensity	Firm	High
Marketing skills	Firm	Low
Mobility of facilities	Firm	Low
Costs/efficiency	Firm	Indeterminate
International information	Firm	Variable by product

A second major problem has to do with determining the long-term purpose of the project. A project may start its life to replace imports in a 'local-for-local' industry, but later be under pressure to start exporting, either because the government's objectives have changed or the industry has become more globally integrated. Initial negotiations typically emphasise how access to the market is to be regulated and the degree of protection offered. Firms may be willing to cede some or all of the ownership of the venture in exchange for market access and some sense of stability in the operating regime. Quite opposite conditions apply for the establishment of 'dependent' exporting in a global industry. Failure to anticipate the possibilities of such a shift in the original deal not only incurs high costs of transition, but may also create conflicting expectations that cannot readily be resolved by subsequent re-negotiation.

The changing nature of the bargaining requirements challenges governments' use of the 'blunt instrument' of general policy, for inflexibility in its application can be counterproductive. Kenya, for example, operates a system of queuing for foreign exchange approvals that slows down decisions and creates uncertainties that impede exports. Malaysia and many other countries have policies of supporting prices for agricultural products, designed among other things to maintain their least efficient farmers. Processing industries using such inputs suffer cost penalties that also impede exports. Further difficulties are created when a shift in international competition upsets the status quo: a firm that operated profitably but inefficiently in a protected local market may find that its very survival is threatened. Even if conflict can be avoided, the government then faces the dilemma of whether it should increase the level of protection, or permit the cheaper, perhaps more technologically advanced, imports to land, or provide adjustment assistance to allow the firm to gain new capabilities.

One of the striking features of the 1980s has been the speed with which many industries have been shifting towards some form of global structure. Equally striking has been the willingness of governments, despite all the obstacles, to change their policy orientation and to move from left to right in figure 1.7. Simultaneous change on both axes can create confusion: the policies adopted are *additive* and force hard choices for governments. None the less, many are adopting much more pragmatic approaches and are becoming more adept in managing selective policies towards those industries and firms that can bring the greatest benefits. For them, as in Brazil, the calculation is not just in terms of today's output but also in terms of gaining greater access to more of the skills needed for tomorrow's competitiveness in an increasingly information- and skill-intensive world.

Managing change of this type requires governments to confront a further issue: taking care to make sure that policy aimed at one of the nine 'cells' in the matrix does not inadvertently wash over to affect behaviour in another. The need to 'ring-fence' agendas can be illustrated by the experience of Malaysia. Ownership restrictions have for long been in place for projects serving the local market, but exporters (mainly in the Export Processing Zones) have been exempt from these provisions, subject to stringent performance targets. When the government introduced its Industrial Coordination Act in the mid-1970s, the exporters feared that the new regulations would apply equally to them. They quickly made it clear that they would reduce future investments and induced the government to revise some of the provisions in the Act and to reaffirm the exemptions.[20]

More generally, the difficulty of keeping agendas separate in this climate of change can be shown by the increasing use of the so-called trade-related investment measures (TRIMS) that link together government incentives and policies for a broad range of trade, investment and financial issues. The emergence of debt-to-equity swaps made contingent on export performance and the debates about how to manage the trade effects of financial institutions are other manifestations of the growing interdependence of previously separate policies. International institutions like the World Bank, IMF, and OECD all have to adjust their policies and reconsider the effectiveness of their loan conditions and structural adjustment programmes.

In a world of increasing divergence among the fortunes of nation states, the role of the international institutions becomes ever more important. Weak institutions will inhibit the development of the kinds of partnership we seek and indeed hope for. How they, too, will adjust to the new challenges will also affect the existing balance in the triangular diplomacy.

LOOKING AHEAD

Change in the international political economy points one way: states are losing power to pursue independent policies and now must master the new game of triangular bargaining. At the same time, the outcomes of the bargaining have been markedly divergent. Indeed, divergence is a *leitmotif* of this book – divergence between continents, between countries in the same continent or region, even within countries over time and among sectors and, finally, divergence between firms – not necessarily in accordance with their national origin. Hardly surprising that, with so many intervening variables, there has been so much variety in the opportunities for change and in the speed with which moves could be made.

The main arguments that follow, therefore, concern the effect of the structural changes on states and firms as interdependencies have multiplied. We aim to show how interdependence affects the options open to both, and why the criteria they use for choice and action can only usefully be considered in the framework first of the triangular nature of the bargaining and second of the matrix of multiple, but related agendas affecting their relationships.

It follows from the above that the skills of managing multiple agendas simultaneously are at a growing premium – for governments, and also for firms. Governments need to manage a series of difficult trade-offs among competing internal and external objectives. They run considerable risk if they fail either to keep some of these agendas separate or to manage the separation in such a way as to appear inconsistent and thus to lose the confidence of prospective investors. Judging where and how to 'ring-fence' the individual agendas stretches the administrative capacity of government. They also run considerable risks if they limit their sights to sectoral preference and fail to perceive that they must also identify strong firms as partners in development.

The more complex the economy, the greater the difficulty in establishing and maintaining explicit policies. Complexity leads to greater reliance on implicit policy, or even in extreme cases to policy setting by default; one can tell only with hindsight what the policy was in the first place. One observer concluded recently that 'the state's most crucial task by far is to provide the rule of law' (Crook, 1989). But the rule of law has to be one that appears reasonably stable and not plagued by frequent, arbitrary alteration.

We also find that many governments adopting more outward-looking, export-promoting policies lack the political will or necessary national consensus to implement them. Central among the many obstacles is that of changing established administrative practice. There can be too many vested interests in the status quo. When change will create too many local 'losers', it will be resisted no matter how strong are the rational arguments. How to address this knotty problem is an issue we pick up at the end of the book, for we believe that all too few observers pay adequate attention to the behavioural aspects of government. The risk that policy pronouncements will not be enacted often lies in the realm of the civil service.

Firms face exactly the same problems of growing complexity. They may seem better equipped for the necessary juggling act because of the greater focus on a limited number of objectives and greater abilities to implement policy in a command structure. Yet even firms are finding that their ability to command from the centre is being challenged by their managers at local levels. Multinationals are far from the monolithic actors they are often deemed to be.

Implicit in these arguments is a general concern about the adequacy of existing theory to explain what has been happening. By working at the micro-level of individual firms and investment projects, we found many instances where we could not have anticipated the outcomes had we predicted them using received wisdom. Our concern is not that the theories are wrong, but incomplete and insufficient to reflect the changes we observe. Because the way the question is asked has much to do with the answer produced, we start the next chapter with a review of theory and a statement of what we see as the critical issues on which progress is needed. Those readers already familiar with the theory may wish to skip over this section; alternatively they may wish to test our ideas against their own.

997

72

WHARTON REPROGRAPHICS

HBR
MARCH–APRIL 1990

The Competitive Advantage of Nations

Michael E. Porter

National prosperity is created, not inherited. It does not grow out of a country's natural endowments, its labor pool, its interest rates, or its currency's value, as classical economics insists.

A nation's competitiveness depends on the capacity of its industry to innovate and upgrade. Companies gain advantage against the world's best competitors because of pressure and challenge. They benefit from having strong domestic rivals, aggressive home-based suppliers, and demanding local customers.

In a world of increasingly global competition, nations have become more, not less, important. As the basis of competition has shifted more and more to the creation and assimilation of knowledge, the role of the nation has grown. Competitive advantage is created and sustained through a highly localized process. Differences in national values, culture, economic structures, institutions, and histories all contribute to competitive success. There are striking differences in the patterns of competitiveness in every country; no nation can or will be competitive in every or even most industries. Ultimately, nations succeed in particular industries because their home environment is the most forward-looking, dynamic, and challenging.

These conclusions, the product of a four-year study of the patterns of competitive success in ten leading trading nations, contradict the conventional wisdom that guides the thinking of many companies and national governments—and that is pervasive today in the United States. (For more about the study, see the insert "Patterns of National Competitive Success.") According to prevailing thinking, labor costs, interest rates, exchange rates, and economies of scale are the most potent determinants of competitiveness. In companies, the words of the day are merger, alliance, strategic partnerships, collaboration, and supranational globalization. Managers are pressing for more government support for particular industries. Among governments, there is a growing tendency to experiment with various policies intended to promote national competitiveness—from efforts to manage exchange rates to new measures to manage trade to policies to relax antitrust—which usually end up only undermining it. (See the insert "What Is National Competitiveness?")

These approaches, now much in favor in both companies and governments, are flawed. They fundamentally misperceive the true sources of competitive advantage. Pursuing them, with all their short-term appeal, will virtually guarantee that the United States—or any other advanced nation—never achieves real and sustainable competitive advantage.

We need a new perspective and new tools—an approach to competitiveness that grows directly out of an analysis of internationally successful industries, without regard for traditional ideology or current intellectual fashion. We need to know, very simply, what works and why. Then we need to apply it.

Harvard Business School professor Michael E. Porter is the author of Competitive Strategy *(Free Press, 1980) and* Competitive Advantage *(Free Press, 1985) and will publish* The Competitive Advantage of Nations *(Free Press) in May 1990.*

Author's note: Michael J. Enright, who served as project coordinator for this study, has contributed valuable suggestions.

Copyright © 1990 by the President and Fellows of Harvard College. All rights reserved.

Patterns of National Competitive Success

To investigate why nations gain competitive advantage in particular industries and the implications for company strategy and national economies, I conducted a four-year study of ten important trading nations: Denmark, Germany, Italy, Japan, Korea, Singapore, Sweden, Switzerland, the United Kingdom, and the United States. I was assisted by a team of more than 30 researchers, most of whom were natives of and based in the nation they studied. The researchers all used the same methodology.

Three nations—the United States, Japan, and Germany—are the world's leading industrial powers. The other nations represent a variety of population sizes, government policies toward industry, social philosophies, geographical sizes, and locations. Together, the ten nations accounted for fully 50% of total world exports in 1985, the base year for statistical analysis.

Most previous analyses of national competitiveness have focused on single nation or bilateral comparisons. By studying nations with widely varying characteristics and circumstances, this study sought to separate the fundamental forces underlying national competitive advantage from the idiosyncratic ones.

In each nation, the study consisted of two parts. The first identified all industries in which the nation's companies were internationally successful, using available statistical data, supplementary published sources, and field interviews. We defined a nation's industry as internationally successful if it *possessed competitive advantage relative to the best worldwide competitors.* Many measures of competitive advantage, such as reported profitability, can be misleading. We chose as the best indicators the presence of substantial and sustained exports to a wide array of other nations and/or significant outbound foreign investment based on skills and assets created in the home country. A nation was considered the home base for a company if it was either a locally owned, indigenous enterprise or managed autonomously although owned by a foreign company or investors. We then created a profile of all the industries in which each nation was internationally successful at three points in time: 1971, 1978, and 1985. The pattern of competitive industries in each economy was far from random: the task was to explain it and how it had changed over time. Of particular interest were the connections or relationships among the nation's competitive industries.

In the second part of the study, we examined the history of competition in particular industries to understand how competitive advantage was created. On the basis of national profiles, we selected over 100 industries or industry groups for detailed study; we examined many more in less detail. We went back as far as necessary to understand how and why the industry began in the nation, how it grew, when and why companies from the nation developed international competitive advantage, and the process by which competitive advantage had been either sustained or lost. The resulting case histories fall short of the work of a good historian in their level of detail, but they do provide insight into the development of both the industry and the nation's economy.

We chose a sample of industries for each nation that represented the most important groups of competitive industries in the economy. The industries studied accounted for a large share of total exports in each nation: more than 20% of total exports in Japan, Germany, and Switzerland, for example, and more than 40% in South Korea. We studied some of the most famous and important international success stories—German high-performance autos and chemicals, Japanese semiconductors and VCRs, Swiss banking and pharmaceuticals, Italian footwear and textiles, U.S. commercial aircraft and motion pictures—and some relatively obscure but highly competitive industries—South Korean pianos, Italian ski boots, and British biscuits. We also added a few industries because they appeared to be paradoxes: Japanese home demand for Western-character typewriters is nearly nonexistent, for example, but Japan holds a strong export and foreign investment position in the industry. We avoided industries that were highly dependent on natural resources: such industries do not form the backbone of advanced economies, and the capacity to compete in them is more explicable using classical theory. We did, however, include a number of more technologically intensive, natural-resource-related industries such as newsprint and agricultural chemicals.

The sample of nations and industries offers a rich empirical foundation for developing and testing the new theory of how countries gain competitive advantage. The accompanying article concentrates on the determinants of competitive advantage in individual industries and also sketches out some of the study's overall implications for government policy and company strategy. A fuller treatment in my book, *The Competitive Advantage of Nations,* develops the theory and its implications in greater depth and provides many additional examples. It also contains detailed descriptions of the nations we studied and the future prospects for their economies.

—Michael E. Porter

How Companies Succeed in International Markets

Around the world, companies that have achieved international leadership employ strategies that differ from each other in every respect. But while every successful company will employ its own particular strategy, the underlying mode of operation—the character and trajectory of all successful companies—is fundamentally the same.

Companies achieve competitive advantage through acts of innovation. They approach innovation in its broadest sense, including both new technologies and new ways of doing things. They perceive a new basis for competing or find better means for competing in old ways. Innovation can be manifested in a new product design, a new production process, a new marketing approach, or a new way of conducting training. Much innovation is mundane and incremental, depending more on a cumulation of small insights and advances than on a single, major technological breakthrough. It often involves ideas that are not even "new"—ideas that have been around, but never vigorously pursued. It always involves investments in skill and knowledge, as well as in physical assets and brand reputations.

Some innovations create competitive advantage by perceiving an entirely new market opportunity or by serving a market segment that others have ignored. When competitors are slow to respond, such innovation yields competitive advantage. For instance, in industries such as autos and home electronics, Japanese companies gained their initial advantage by emphasizing smaller, more compact, lower capacity models that foreign competitors disdained as less profitable, less important, and less attractive.

In international markets, innovations that yield competitive advantage anticipate both domestic and foreign needs. For example, as international concern for product safety has grown, Swedish companies like Volvo, Atlas Copco, and AGA have succeeded by anticipating the market opportunity in this area. On the other hand, innovations that respond to concerns or circumstances that are peculiar to the home market can actually retard international competitive success. The lure of the huge U.S. defense market, for instance, has diverted the attention of U.S. materials and machine-tool companies from attractive, global commercial markets.

Information plays a large role in the process of innovation and improvement—information that either is not available to competitors or that they do not seek. Sometimes it comes from simple investment in research and development or market research; more often, it comes from effort and from openness and from looking in the right place unencumbered by blinding assumptions or conventional wisdom.

This is why innovators are often outsiders from a different industry or a different country. Innovation may come from a new company, whose founder has a nontraditional background or was simply not appreciated in an older, established company. Or the capacity for innovation may come into an existing company through senior managers who are new to the particular industry and thus more able to perceive opportunities and more likely to pursue them. Or innovation may occur as a company diversifies, bringing new resources, skills, or perspectives to another industry. Or innovations may come from another nation with different circumstances or different ways of competing.

With few exceptions, innovation is the result of unusual effort. The company that successfully implements a new or better way of competing pursues its approach with dogged determination, often in the face of harsh criticism and tough obstacles. In fact, to succeed, innovation usually requires pressure, necessity, and even adversity: the fear of loss often proves more powerful than the hope of gain.

Once a company achieves competitive advantage through an innovation, it can sustain it only through relentless improvement. Almost any advantage can be imitated. Korean companies have already matched the ability of their Japanese rivals to mass-produce standard color televisions and VCRs; Brazilian companies have assembled technology and designs comparable to Italian competitors in casual leather footwear.

Competitors will eventually and inevitably overtake any company that stops improving and innovating. Sometimes early-mover advantages such as customer relationships, scale economies in existing technologies, or the loyalty of distribution channels are enough to permit a stagnant company to retain its entrenched position for years or even decades. But sooner or later, more dynamic rivals will find a way to innovate around these advantages or create a better or cheaper way of doing things. Italian appliance producers, which competed successfully on the basis of cost in selling midsize and compact appliances through large retail chains, rested too long on this initial advantage. By developing more differentiated products and creating strong brand franchises, German competitors have begun to gain ground.

Ultimately, the only way to sustain a competitive advantage is to *upgrade it*—to move to more sophisticated types. This is precisely what Japanese automakers have done. They initially penetrated foreign markets with small, inexpensive compact cars of adequate quality and competed on the basis of lower labor costs. Even while their labor-cost advantage persisted, however, the Japanese companies were up-

What Is National Competitiveness?

National competitiveness has become one of the central preoccupations of government and industry in every nation. Yet for all the discussion, debate, and writing on the topic, there is still no persuasive theory to explain national competitiveness. What is more, there is not even an accepted definition of the term "competitiveness" as applied to a nation. While the notion of a competitive company is clear, the notion of a competitive nation is not.

Some see national competitiveness as a macroeconomic phenomenon, driven by variables such as exchange rates, interest rates, and government deficits. But Japan, Italy, and South Korea have all enjoyed rapidly rising living standards despite budget deficits; Germany and Switzerland despite appreciating currencies; and Italy and Korea despite high interest rates.

Others argue that competitiveness is a function of cheap and abundant labor. But Germany, Switzerland, and Sweden have all prospered even with high wages and labor shortages. Besides, shouldn't a nation seek higher wages for its workers as a goal of competitiveness?

Another view connects competitiveness with bountiful natural resources. But how, then, can one explain the success of Germany, Japan, Switzerland, Italy, and South Korea—countries with limited natural resources?

More recently, the argument has gained favor that competitiveness is driven by government policy: targeting, protection, import promotion, and subsidies have propelled Japanese and South Korean auto, steel, shipbuilding, and semiconductor industries into global preeminence. But a closer look reveals a spotty record. In Italy, government intervention has been ineffectual—but Italy has experienced a boom in world export share second only to Japan. In Germany, direct government intervention in exporting industries is rare. And even in Japan and South Korea, government's role in such important industries as facsimile machines, copiers, robotics, and advanced materials has been modest; some of the most frequently cited examples, such as sewing machines, steel, and shipbuilding, are now quite dated.

A final popular explanation for national competitiveness is differences in management practices, including management-labor relations. The problem here, however, is that different industries require different approaches to management. The successful management practices governing small, private, and loosely organized Italian family companies in footwear, textiles, and jewelry, for example, would produce a management disaster if applied to German chemical or auto companies, Swiss pharmaceutical makers, or American aircraft producers. Nor is it possible to generalize about management-labor relations. Despite the commonly held view that powerful unions undermine competitive advantage, unions are strong in Germany and Sweden—and both countries boast internationally preeminent companies.

Clearly, none of these explanations is fully satisfactory; none is sufficient by itself to rationalize the competitive position of industries within a national border. Each contains some truth; but a broader, more complex set of forces seems to be at work.

The lack of a clear explanation signals an even more fundamental question. What is a "competitive" nation in the first place? Is a "competitive" nation one where every company or industry is competitive? No nation meets this test. Even Japan has large sectors of its economy that fall far behind the world's best competitors.

Is a "competitive" nation one whose exchange rate makes its goods price competitive in international markets? Both Germany and Japan have enjoyed remarkable gains in their standards of living—and experienced sustained periods of strong currency and rising prices. Is a "competitive" nation one with a large positive balance of trade? Switzerland has roughly balanced trade; Italy has a chronic trade deficit—both nations enjoy strongly rising national income. Is a "competitive" nation one with low labor costs? India and Mexico both have low wages and low labor costs—but neither seems an attractive industrial model.

The only meaningful concept of competitiveness at the national level is *productivity*. The principal goal of a nation is to produce a high and rising standard of living for its citizens. The ability to do so depends on the productivity with which a nation's labor and capital are employed. Productivity is the value of the output produced by a unit of labor or capital. Productivity depends on both the quality and features of products (which determine the prices that they can command) and the efficiency with which they are produced. Productivity is the prime determinant of a nation's long-run standard of living; it is the root cause of national per capita income. The productivity of human resources determines employee wages; the productivity with which capital is employed determines the return it earns for its holders.

A nation's standard of living depends on the capacity of its companies to achieve high levels of productivity—and to increase productivity over time. Sustained productivity growth requires that an economy continually *upgrade itself.* A nation's companies must relentlessly improve productivity in existing industries by raising product quality, adding desirable features, improving product technology, or boosting production efficiency. They must develop the necessary capabilities to compete in more and more sophisticated industry segments, where productivity is generally high. They must finally develop the capability to compete in entirely new, sophisticated industries.

International trade and foreign investment can both improve a nation's productivity as well as threaten it. They support rising national productivity by allowing a nation to specialize in those industries and segments of industries where its companies are more productive and to import where its companies are less productive. No nation can be competitive in everything. The ideal is to deploy the nation's limited pool of human and other resources into the most productive uses. Even those nations with the highest standards of living have many industries in which local companies are uncompetitive.

Yet international trade and foreign investment also can threaten productivity growth. They expose a nation's industries to the test of international standards of productivity. An industry will lose out if its productivity is not sufficiently higher than foreign rivals' to offset any advantages in local wage rates. If a nation loses the ability to compete in a range of high-productivity/high-wage industries, its standard of living is threatened.

Defining national competitiveness as achieving a trade surplus or balanced trade per se is inappropriate. The expansion of exports because of low wages and a weak currency, at the same time that the nation imports sophisticated goods that its companies cannot produce competitively, may bring trade into balance or surplus but lowers the nation's standard of living. Competitiveness also does not mean jobs. It's the *type* of jobs, not just the ability to employ citizens at low wages, that is decisive for economic prosperity.

Seeking to explain "competitiveness" at the national level, then, is to answer the wrong question. What we must understand instead is the determinants of productivity and the rate of productivity growth. To find answers, we must focus not on the economy as a whole but on *specific industries and industry segments*. We must understand how and why commercially viable skills and technology are created, which can only be fully understood at the level of particular industries. It is the outcome of the thousands of struggles for competitive advantage against foreign rivals in particular segments and industries, in which products and processes are created and improved, that underpins the process of upgrading national productivity.

When one looks closely at any national economy, there are striking differences among a nation's industries in competitive success. International advantage is often concentrated in particular industry segments. German exports of cars are heavily skewed toward high-performance cars, while Korean exports are all compacts and subcompacts. In many industries and segments of industries, the competitors with true international competitive advantage are *based in only a few nations*.

Our search, then, is for the decisive characteristic of a nation that allows its companies to create and sustain competitive advantage in particular fields—the search is for the competitive advantage of nations. We are particularly concerned with the determinants of international success in technology- and skill-intensive segments and industries, which underpin high and rising productivity.

Classical theory explains the success of nations in particular industries based on so-called factors of production such as land, labor, and natural resources. Nations gain factor-based comparative advantage in industries that make intensive use of the factors they possess in abundance. Classical theory, however, has been overshadowed in advanced industries and economies by the globalization of competition and the power of technology.

A new theory must recognize that in modern international competition, companies compete with global strategies involving not only trade but also foreign investment. What a new theory must explain is why a nation provides a favorable *home base* for companies that compete internationally. The home base is the nation in which the essential competitive advantages of the enterprise are created and sustained. It is where a company's strategy is set, where the core product and process technology is created and maintained, and where the most productive jobs and most advanced skills are located. The presence of the home base in a nation has the greatest positive influence on other linked domestic industries and leads to other benefits in the nation's economy. While the ownership of the company is often concentrated at the home base, the nationality of shareholders is secondary.

A new theory must move beyond comparative advantage to the competitive advantage of a nation. It must reflect a rich conception of competition that includes segmented markets, differentiated products, technology differences, and economies of scale. A new theory must go beyond cost and explain why companies from some nations are better than others at creating advantages based on quality, features, and new product innovation. A new theory must begin from the premise that competition is dynamic and evolving; it must answer the questions: Why do some companies based in some nations innovate more than others? Why do some nations provide an environment that enables companies to improve and innovate faster than foreign rivals?

—Michael E. Porter

grading. They invested aggressively to build large modern plants to reap economies of scale. Then they became innovators in process technology, pioneering just-in-time production and a host of other quality and productivity practices. These process improvements led to better product quality, better repair records, and better customer-satisfaction ratings than foreign competitors had. Most recently, Japanese automakers have advanced to the vanguard of product technology and are introducing new, premium brand names to compete with the world's most prestigious passenger cars.

The example of the Japanese automakers also illustrates two additional prerequisites for sustaining competitive advantage. First, a company must adopt a global approach to strategy. It must sell its product worldwide, under its own brand name, through international marketing channels that it controls. A truly global approach may even require the company to locate production or R&D facilities in other nations to take advantage of lower wage rates, to gain or improve market access, or to take advantage of foreign technology. Second, creating more sustainable advantages often means that a company must make its existing advantage obsolete—even while it is still an advantage. Japanese auto companies recognized this; either they would make their advantage obsolete, or a competitor would do it for them.

As this example suggests, innovation and change are inextricably tied together. But change is an unnatural act, particularly in successful companies; powerful forces are at work to avoid and defeat it. Past approaches become institutionalized in standard operating procedures and management controls. Training emphasizes the one correct way to do anything; the construction of specialized, dedicated facilities solidifies past practice into expensive brick and mortar; the existing strategy takes on an aura of invincibility and becomes rooted in the company culture.

Successful companies tend to develop a bias for predictability and stability; they work on defending what they have. Change is tempered by the fear that there is much to lose. The organization at all levels filters out information that would suggest new approaches, modifications, or departures from the norm. The internal environment operates like an immune system to isolate or expel "hostile" individuals who challenge current directions or established thinking. Innovation ceases; the company becomes stagnant; it is only a matter of time before aggressive competitors overtake it.

The Diamond of National Advantage

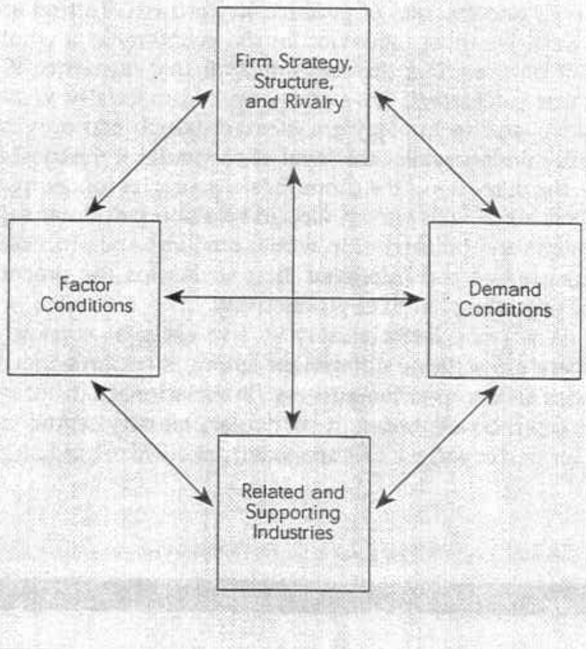
Why are certain companies based in certain nations capable of consistent innovation? Why do they ruthlessly pursue improvements, seeking an ever-more sophisticated source of competitive advantage? Why are they able to overcome the substantial barriers to change and innovation that so often accompany success?

The answer lies in four broad attributes of a nation, attributes that individually and as a system constitute the diamond of national advantage, the playing field that each nation establishes and operates for its industries. These attributes are.

1. *Factor Conditions.* The nation's position in factors of production, such as skilled labor or infrastructure, necessary to compete in a given industry.
2. *Demand Conditions.* The nature of home-market demand for the industry's product or service.
3. *Related and Supporting Industries.* The presence or absence in the nation of supplier industries and other related industries that are internationally competitive.
4. *Firm Strategy, Structure, and Rivalry.* The conditions in the nation governing how companies are created, organized, and managed, as well as the nature of domestic rivalry.

These determinants create the national environment in which companies are born and learn how to compete. (See the diagram "Determinants of National Competitive Advantage.") Each point on the

Determinants of National Competitive Advantage


Firm Strategy, Structure, and Rivalry
Factor Conditions
Demand Conditions
Related and Supporting Industries

diamond—and the diamond as a system—affects essential ingredients for achieving international competitive success: the availability of resources and skills necessary for competitive advantage in an industry; the information that shapes the opportunities that companies perceive and the directions in which they deploy their resources and skills; the goals of the owners, managers, and individuals in companies; and most important, the pressures on companies to invest and innovate. (See the insert "How the Diamond Works: The Italian Ceramic Tile Industry.")

When a national environment permits and supports the most rapid accumulation of specialized assets and skills—sometimes simply because of greater effort and commitment—companies gain a competitive advantage. When a national environment affords better ongoing information and insight into product and process needs, companies gain a competitive advantage. Finally, when the national environment pressures companies to innovate and invest, companies both gain a competitive advantage and upgrade those advantages over time.

Factor Conditions. According to standard economic theory, factors of production—labor, land, natural resources, capital, infrastructure—will determine the flow of trade. A nation will export those goods that make most use of the factors with which it is relatively well endowed. This doctrine, whose origins date back to Adam Smith and David Ricardo and that is embedded in classical economics, is at best incomplete and at worst incorrect.

In the sophisticated industries that form the backbone of any advanced economy, a nation does not inherit but instead creates the most important factors of production—such as skilled human resources or a scientific base. Moreover, the stock of factors that a nation enjoys at a particular time is less important than the rate and efficiency with which it creates, upgrades, and deploys them in particular industries.

The most important factors of production are those that involve sustained and heavy investment and are specialized. Basic factors, such as a pool of labor or a local raw-material source, do not constitute an advantage in knowledge-intensive industries. Companies can access them easily through a global strategy or circumvent them through technology. Contrary to conventional wisdom, simply having a general work force that is high school or even college educated represents no competitive advantage in modern international competition. To support competitive advantage, a factor must be highly specialized to an industry's particular needs—a scientific institute specialized in optics, a pool of venture capital to fund software companies. These factors are more scarce, more difficult for foreign competitors to imitate—and they require sustained investment to create.

Nations succeed in industries where they are particularly good at factor creation. Competitive advantage results from the presence of world-class institutions that first create specialized factors and then continually work to upgrade them. Denmark has two hospitals that concentrate in studying and treating diabetes—and a world-leading export position in insulin. Holland has premier research institutes in the cultivation, packaging, and shipping of flowers, where it is the world's export leader.

What is not so obvious, however, is that selective disadvantages in the more basic factors can prod a company to innovate and upgrade—a disadvantage in a static model of competition can become an advantage in a dynamic one. When there is an ample supply of cheap raw materials or abundant labor, companies can simply rest on these advantages and often deploy them inefficiently. But when companies face a selective disadvantage, like high land costs, labor shortages, or the lack of local raw materials, they *must* innovate and upgrade to compete.

Implicit in the oft-repeated Japanese statement, "We are an island nation with no natural resources," is the understanding that these deficiencies have only served to spur Japan's competitive innovation. Just-in-time production, for example, economized on prohibitively expensive space. Italian steel producers in the Brescia area faced a similar set of disadvantages: high capital costs, high energy costs, and no local raw materials. Located in Northern Lombardy, these privately owned companies faced staggering logistics costs due to their distance from southern ports and the inefficiencies of the state-owned Italian transportation system. The result: they pioneered technologically advanced minimills that require only modest capital investment, use less energy, employ scrap metal as the feedstock, are efficient at small scale, and permit producers to locate close to sources of scrap and end-use customers. In other words, they converted factor disadvantages into competitive advantage.

Disadvantages can become advantages only under certain conditions. First, they must send companies proper signals about circumstances that will spread to other nations, thereby equipping them to innovate in advance of foreign rivals. Switzerland, the nation that experienced the first labor shortages after World War II, is a case in point. Swiss companies responded to the disadvantage by upgrading labor productivity and seeking higher value, more sustainable market segments. Companies in most other parts of the world, where there were still ample workers, focused their attention on other issues, which resulted in slower upgrading.

The second condition for transforming disadvan-

How the Diamond Works: The Italian Ceramic Tile Industry

In 1987, Italian companies were world leaders in the production and export of ceramic tiles, a $10 billion industry. Italian producers, concentrated in and around the small town of Sassuolo in the Emilia-Romagna region, accounted for about 30% of world production and almost 60% of world exports. The Italian trade surplus that year in ceramic tiles was about $1.4 billion.

The development of the Italian ceramic tile industry's competitive advantage illustrates how the diamond of national advantage works. Sassuolo's sustainable competitive advantage in ceramic tiles grew not from any static or historical advantage but from dynamism and change. Sophisticated and demanding local buyers, strong and unique distribution channels, and intense rivalry among local companies created constant pressure for innovation. Knowledge grew quickly from continuous experimentation and cumulative production experience. Private ownership of the companies and loyalty to the community spawned intense commitment to invest in the industry.

Tile producers benefited as well from a highly developed set of local machinery suppliers and other supporting industries, producing materials, services, and infrastructure. The presence of world-class, Italian-related industries also reinforced Italian strength in tiles. Finally, the geographic concentration of the entire cluster supercharged the whole process. Today foreign companies compete against an entire subculture. The organic nature of this system represents the most sustainable advantage of Sassuolo's ceramic tile companies.

The Origins of the Italian Industry

Tile production in Sassuolo grew out of the earthenware and crockery industry, whose history traces back to the thirteenth century. Immediately after World War II, there were only a handful of ceramic tile manufacturers in and around Sassuolo, all serving the local market exclusively.

Demand for ceramic tiles within Italy began to grow dramatically in the immediate postwar years, as the reconstruction of Italy triggered a boom in building materials of all kinds. Italian demand for ceramic tiles was particularly great due to the climate, local tastes, and building techniques.

Because Sassuolo was in a relatively prosperous part of Italy, there were many who could combine the modest amount of capital and necessary organizational skills to start a tile company. In 1955, there were 14 Sassuolo area tile companies; by 1962, there were 102.

The new tile companies benefited from a local pool of mechanically trained workers. The region around Sassuolo was home to Ferrari, Maserati, Lamborghini, and other technically sophisticated companies. As the tile industry began to grow and prosper, many engineers and skilled workers gravitated to the successful companies.

The Emerging Italian Tile Cluster

Initially, Italian tile producers were dependent on foreign sources of raw materials and production technology. In the 1950s, the principal raw materials used to make tiles were kaolin (white) clays. Since there were red- but no white-clay deposits near Sassuolo, Italian producers had to import the clays from the United Kingdom. Tile-making equipment was also imported in the 1950s and 1960s: kilns from Germany, America, and France; presses for forming tiles from Germany. Sassuolo tile makers had to import even simple glazing machines.

Over time, the Italian tile producers learned how to modify imported equipment to fit local circumstances: red versus white clays, natural gas versus heavy oil. As process technicians from tile companies left to start their own equipment companies, a local machinery industry arose in Sassuolo. By 1970, Italian companies had emerged as world-class producers of kilns and presses; the earlier situation had exactly reversed: they were exporting their red-clay equipment for foreigners to use with white clays.

The relationship between Italian tile and equipment manufacturers was a mutually supporting one, made even more so by close proximity. In the mid-1980s, there were some 200 Italian equipment manufacturers; more than 60% were located in the Sassuolo area. The equipment manufacturers competed fiercely for local business, and tile manufacturers benefited from better prices and more advanced equipment than their foreign rivals.

As the emerging tile cluster grew and concentrated in the Sassuolo region, a pool of skilled workers and technicians developed, including engineers, production specialists, maintenance workers, service technicians, and design personnel. The industry's geographic concentration encouraged other supporting companies to form, offering molds, packaging materials, glazes, and transportation services. An array of small, specialized consulting companies emerged to give advice to tile producers on plant design, logistics, and commercial, advertising, and fiscal matters.

With its membership concentrated in the Sassuolo area, Assopiastrelle, the ceramic tile industry association, began offering services in areas of common interest: bulk purchasing, foreign-market research, and consulting on fiscal and legal matters. The growing tile cluster stimulated the formation of a new, specialized factor-creating institution: in 1976, a consortium of the

University of Bologna, regional agencies, and the ceramic industry association founded the Centro Ceramico di Bologna, which conducted process research and product analysis.

Sophisticated Home Demand

By the mid-1960s, per-capita tile consumption in Italy was considerably higher than in the rest of the world. The Italian market was also the world's most sophisticated. Italian customers, who were generally the first to adopt new designs and features, and Italian producers, who constantly innovated to improve manufacturing methods and create new designs, progressed in a mutually reinforcing process.

The uniquely sophisticated character of domestic demand also extended to retail outlets. In the 1960s, specialized tile showrooms began opening in Italy. By 1985, there were roughly 7,600 specialized showrooms handling approximately 80% of domestic sales, far more than in other nations. In 1976, the Italian company Piemme introduced tiles by famous designers to gain distribution outlets and to build brand name awareness among consumers. This innovation drew on another related industry, design services, in which Italy was world leader, with over $10 billion in exports.

Sassuolo Rivalry

The sheer number of tile companies in the Sassuolo area created intense rivalry. News of product and process innovations spread rapidly, and companies seeking technological, design, and distribution leadership had to improve constantly.

Proximity added a personal note to the intense rivalry. All of the producers were privately held, most were family run. The owners all lived in the same area, knew each other, and were the leading citizens of the same towns.

Pressures to Upgrade

In the early 1970s, faced with intense domestic rivalry, pressure from retail customers, and the shock of the 1973 energy crisis, Italian tile companies struggled to reduce gas and labor costs. These efforts led to a technological breakthrough, the rapid single-firing process, in which the hardening process, material transformation, and glaze-fixing all occurred in one pass through the kiln. A process that took 225 employees using the double-firing method needed only 90 employees using single-firing roller kilns. Cycle time dropped from 16 to 20 hours to only 50 to 55 minutes.

The new, smaller, and lighter equipment was also easier to export. By the early 1980s, exports from Italian equipment manufacturers exceeded domestic sales; in 1988, exports represented almost 80% of total sales.

Working together, tile manufacturers and equipment manufacturers made the next important breakthrough during the mid- and late 1970s: the development of materials-handling equipment that transformed tile manufacture from a batch process to a continuous process. The innovation reduced high labor costs—which had been a substantial selective factor disadvantage facing Italian tile manufacturers.

The common perception is that Italian labor costs were lower during this period than those in the United States and Germany. In those two countries, however, different jobs had widely different wages. In Italy, wages for different skill categories were compressed, and work rules constrained manufacturers from using overtime or multiple shifts. The restriction proved costly: once cool, kilns are expensive to reheat and are best run continuously. Because of this factor disadvantage, the Italian companies were the first to develop continuous, automated production.

Internationalization

By 1970, Italian domestic demand had matured. The stagnant Italian market led companies to step up their efforts to pursue foreign markets. The presence of related and supporting Italian industries helped in the export drive. Individual tile manufacturers began advertising in Italian and foreign home-design and architectural magazines, publications with wide global circulation among architects, designers, and consumers. This heightened awareness reinforced the quality image of Italian tiles. Tile makers were also able to capitalize on Italy's leading world export positions in related industries like marble, building stone, sinks, washbasins, furniture, lamps, and home appliances.

Assopiastrelle, the industry association, established trade-promotion offices in the United States in 1980, in Germany in 1984, and in France in 1987. It organized elaborate trade shows in cities ranging from Bologna to Miami and ran sophisticated advertising. Between 1980 and 1987, the association spent roughly $8 million to promote Italian tiles in the United States.

—Michael J. Enright and Paolo Tenti

Michael J. Enright, a doctoral student in business economics at the Harvard Business School, performed numerous research and supervisory tasks for The Competitive Advantage of Nations. *Paolo Tenti was responsible for the Italian part of research undertaken for the book. He is a consultant in strategy and finance for Monitor Company and Analysis F.A.–Milan.*

tages into advantages is favorable circumstances elsewhere in the diamond—a consideration that applies to almost all determinants. To innovate, companies must have access to people with appropriate skills and have home-demand conditions that send the right signals. They must also have active domestic rivals who create pressure to innovate. Another precondition is company goals that lead to sustained commitment to the industry. Without such a commitment and the presence of active rivalry, a company may take an easy way around a disadvantage rather than using it as a spur to innovation.

For example, U.S. consumer-electronics companies, faced with high relative labor costs, chose to leave the product and production process largely unchanged and move labor-intensive activities to Taiwan and other Asian countries. Instead of upgrading their sources of advantage, they settled for labor-cost parity. On the other hand, Japanese rivals, confronted with intense domestic competition and a mature home market, chose to eliminate labor through automation. This led to lower assembly costs, to products with fewer components and to improved quality and reliability. Soon Japanese companies were building assembly plants in the United States—the place U.S. companies had fled.

Demand Conditions. It might seem that the globalization of competition would diminish the importance of home demand. In practice, however, this is simply not the case. In fact, the composition and character of the home market usually has a disproportionate effect on how companies perceive, interpret, and respond to buyer needs. Nations gain competitive advantage in industries where the home demand gives their companies a clearer or earlier picture of emerging buyer needs, and where demanding buyers pressure companies to innovate faster and achieve more sophisticated competitive advantages than their foreign rivals. The size of home demand proves far less significant than the character of home demand.

Home-demand conditions help build competitive advantage when a particular industry segment is larger or more visible in the domestic market than in foreign markets. The larger market segments in a nation receive the most attention from the nation's companies; companies accord smaller or less desirable segments a lower priority. A good example is hydraulic excavators, which represent the most widely used type of construction equipment in the Japanese domestic market—but which comprise a far smaller proportion of the market in other advanced nations. This segment is one of the few where there are vigorous Japanese international competitors and where Caterpillar does not hold a substantial share of the world market.

More important than the mix of segments per se is the nature of domestic buyers. A nation's companies gain competitive advantage if domestic buyers are the world's most sophisticated and demanding buyers for the product or service. Sophisticated, demanding buyers provide a window into advanced customer needs; they pressure companies to meet high standards; they prod them to improve, to innovate, and to upgrade into more advanced segments. As with factor conditions, demand conditions provide advantages by forcing companies to respond to tough challenges.

Especially stringent needs arise because of local values and circumstances. For example, Japanese consumers, who live in small, tightly packed homes, must contend with hot, humid summers and high-cost electrical energy—a daunting combination of circumstances. In response, Japanese companies have pioneered compact, quiet air-conditioning units powered by energy-saving rotary compressors. In industry after industry, the tightly constrained requirements of the Japanese market have forced companies to innovate, yielding products that are *kei-haku-tan-sho*—light, thin, short, small—and that are internationally accepted.

Local buyers can help a nation's companies gain advantage if their needs anticipate or even shape those of other nations—if their needs provide ongoing "early-warning indicators" of global market trends. Sometimes anticipatory needs emerge because a nation's political values foreshadow needs that will grow elsewhere. Sweden's long-standing concern for handicapped people has spawned an increasingly competitive industry focused on special needs. Denmark's environmentalism has led to success for companies in water-pollution control equipment and windmills.

More generally, a nation's companies can anticipate global trends if the nation's values are spreading—that is, if the country is exporting its values and tastes as well as its products. The international success of U.S. companies in fast food and credit cards, for example, reflects not only the American desire for convenience but also the spread of these tastes to the rest of the world. Nations export their values and tastes through media, through training foreigners, through political influence, and through the foreign activities of their citizens and companies.

Related and Supporting Industries. The third broad determinant of national advantage is the presence in the nation of related and supporting industries that are internationally competitive. Internationally competitive home-based suppliers create advantages in downstream industries in several ways. First, they deliver the most cost-effective inputs in an efficient, early, rapid, and sometimes preferential way. Italian gold and silver jewelry companies lead the world in that industry in part because other Italian companies supply two-thirds of the world's jewelry-making and precious-metal recycling machinery.

Far more significant than mere access to components and machinery, however, is the advantage that home-based related and supporting industries provide in innovation and upgrading—an advantage based on close working relationships. Suppliers and end-users located near each other can take advantage of short lines of communication, quick and constant flow of information, and an ongoing exchange of ideas and innovations. Companies have the opportunity to influence their suppliers' technical efforts and can serve as test sites for R&D work, accelerating the pace of innovation.

The illustration of "The Italian Footwear Cluster" offers a graphic example of how a group of close-by, supporting industries creates competitive advantage in a range of interconnected industries that are all internationally competitive. Shoe producers, for instance, interact regularly with leather manufacturers on new styles and manufacturing techniques and learn about new textures and colors of leather when they are still on the drawing boards. Leather manufacturers gain early insights into fashion trends, helping them to plan new products. The interaction is mutually advantageous and self-reinforcing, but it does not happen automatically: it is helped by proximity, but occurs only because companies and suppliers work at it.

The nation's companies benefit most when the suppliers are, themselves, global competitors. It is ultimately self-defeating for a company or country to create "captive" suppliers who are totally dependent on the domestic industry and prevented from serving foreign competitors. By the same token, a nation need not be competitive in all supplier industries for its companies to gain competitive advantage. Companies can readily source from abroad materials, components, or technologies without a major effect on innovation or performance of the industry's products. The same is true of other generalized technologies—like electronics or software—where the industry represents a narrow application area.

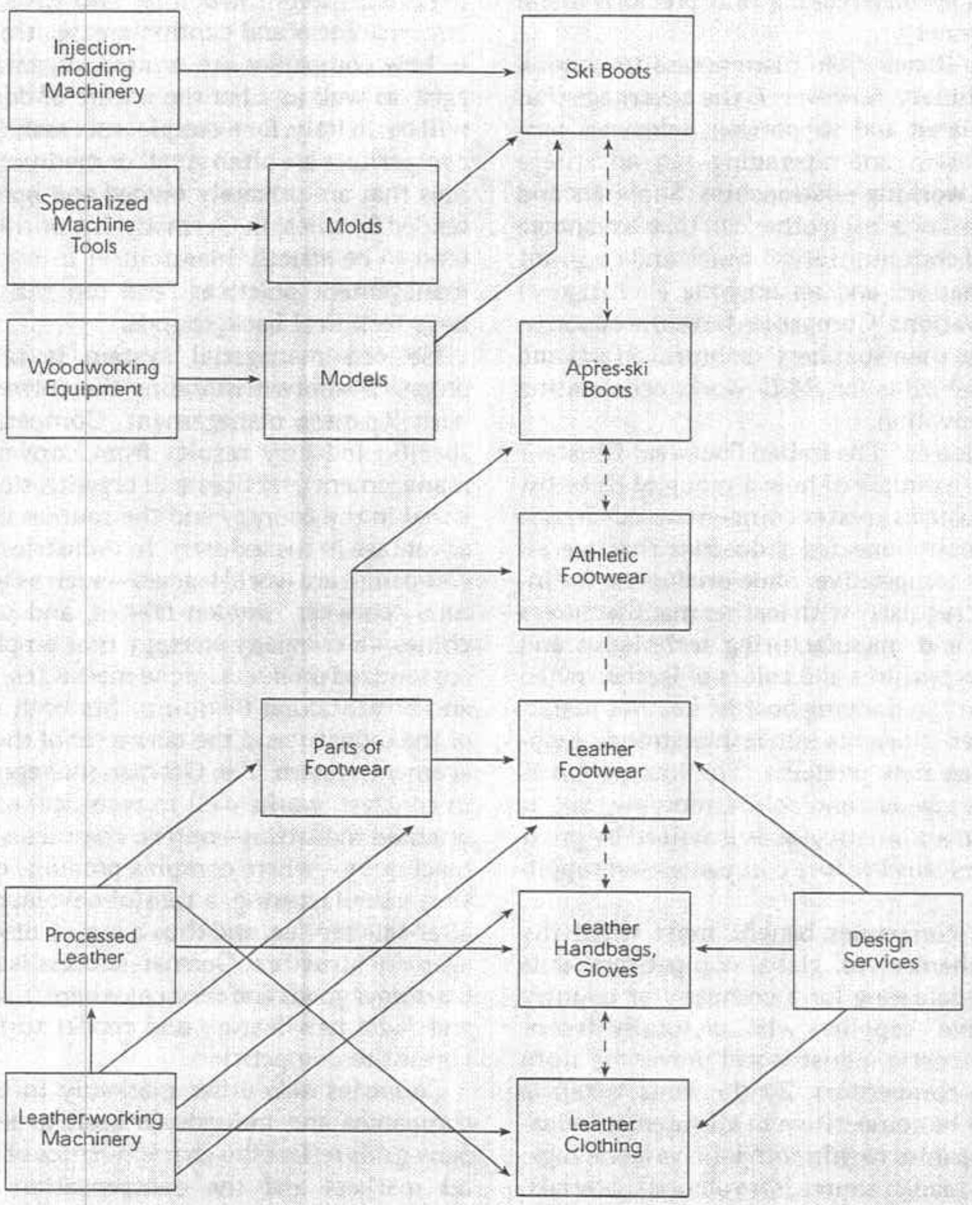
Home-based competitiveness in related industries provides similar benefits: information flow and technical interchange speed the rate of innovation and upgrading. A home-based related industry also increases the likelihood that companies will embrace new skills, and it also provides a source of entrants who will bring a novel approach to competing. The Swiss success in pharmaceuticals emerged out of previous international success in the dye industry, for example; Japanese dominance in electronic musical keyboards grows out of success in acoustic instruments combined with a strong position in consumer electronics.

Firm Strategy, Structure, and Rivalry. National circumstances and context create strong tendencies in how companies are created, organized, and managed, as well as what the nature of domestic rivalry will be. In Italy, for example, successful international competitors are often small or medium-sized companies that are privately owned and operated like extended families; in Germany, in contrast, companies tend to be strictly hierarchical in organization and management practices, and top managers usually have technical backgrounds.

No one managerial system is universally appropriate—notwithstanding the current fascination with Japanese management. Competitiveness in a specific industry results from convergence of the management practices and organizational modes favored in the country and the sources of competitive advantage in the industry. In industries where Italian companies are world leaders—such as lighting, furniture, footwear, woolen fabrics, and packaging machines—a company strategy that emphasizes focus, customized products, niche marketing, rapid change, and breathtaking flexibility fits both the dynamics of the industry and the character of the Italian management system. The German management system, in contrast, works well in technical or engineering-oriented industries—optics, chemicals, complicated machinery—where complex products demand precision manufacturing, a careful development process, after-sale service, and thus a highly disciplined management structure. German success is much rarer in consumer goods and services where image marketing and rapid new-feature and model turnover are important to competition.

Countries also differ markedly in the goals that companies and individuals seek to achieve. Company goals reflect the characteristics of national capital markets and the compensation practices for managers. For example, in Germany and Switzerland, where banks comprise a substantial part of the nation's shareholders, most shares are held for long-term appreciation and are rarely traded. Companies do well in mature industries, where ongoing investment in R&D and new facilities is essential but returns may be only moderate. The United States is at the opposite extreme, with a large pool of risk capital but widespread trading of public companies and a strong emphasis by investors on quarterly and annual share-price appreciation. Management compensation is heavily based on annual bonuses tied to individual results. America does well in relatively new industries, like software and biotechnology, or ones where equity funding of new companies feeds

The Italian Footwear Cluster


Injection-molding Machinery
Ski Boots
Specialized Machine Tools
Molds
Woodworking Equipment
Models
Apres-ski Boots
Athletic Footwear
Parts of Footwear
Leather Footwear
Processed Leather
Leather Handbags, Gloves
Design Services
Leather-working Machinery
Leather Clothing

active domestic rivalry, like specialty electronics and services. Strong pressures leading to underinvestment, however, plague more mature industries.

Individual motivation to work and expand skills is also important to competitive advantage. Outstanding talent is a scarce resource in any nation. A nation's success largely depends on the types of education its talented people choose, where they choose to work, and their commitment and effort. The goals a nation's institutions and values set for individuals and companies, and the prestige it attaches to certain industries, guide the flow of capital and human resources—which, in turn, directly affects the competitive performance of certain industries. Nations tend to be competitive in activities that people admire or depend on—the activities from which the nation's heroes emerge. In Switzerland, it is banking and pharmaceuticals. In Israel, the highest callings have been agriculture and defense-related fields. Sometimes it is hard to distinguish between cause and effect. Attaining international success can make an industry prestigious, reinforcing its advantage.

The presence of strong local rivals is a final, and powerful, stimulus to the creation and persistence of competitive advantage. This is true of small countries, like Switzerland, where the rivalry among its pharmaceutical companies, Hoffmann-La Roche, Ciba-Geigy, and Sandoz, contributes to a leading worldwide position. It is true in the United States in the computer and software industries. Nowhere is the role of fierce rivalry more apparent than in Japan, where there are 112 companies competing in machine tools, 34 in semiconductors, 25 in audio equipment, 15 in cameras—in fact, there are usually double figures in the industries in which Japan boasts global dominance. (See the table "Estimated Number of Japanese Rivals in Selected Industries.") Among all the points on the diamond, domestic rivalry is arguably the most important because of the powerfully stimulating effect it has on all the others.

Conventional wisdom argues that domestic competition is wasteful: it leads to duplication of effort and prevents companies from achieving economies of scale. The "right solution" is to embrace one or two national champions, companies with the scale and strength to tackle foreign competitors, and to guarantee them the necessary resources, with the government's blessing. In fact, however, most national champions are uncompetitive, although heavily subsidized and protected by their government. In many of the prominent industries in which there is only one national rival, such as aerospace and telecommunications, government has played a large role in distorting competition.

Static efficiency is much less important than dynamic improvement, which domestic rivalry uniquely spurs. Domestic rivalry, like any rivalry, creates pressure on companies to innovate and improve. Local rivals push each other to lower costs, improve quality and service, and create new products and processes. But unlike rivalries with foreign competitors, which tend to be analytical and distant, local rivalries often go beyond pure economic or business competition and become intensely personal. Domestic rivals engage in active feuds; they compete not only for market share but also for people, for technical excellence, and perhaps most important, for "bragging rights." One domestic rival's success proves to others that advancement is possible and often attracts new rivals to the industry. Companies often attribute the success of foreign rivals to "unfair" advantages. With domestic rivals, there are no excuses.

Geographic concentration magnifies the power of domestic rivalry. This pattern is strikingly common around the world: Italian jewelry companies are located around two towns, Arezzo and Valenza Po; cutlery companies in Solingen, West Germany and Seki, Japan; pharmaceutical companies in Basel, Switzerland; motorcycles and musical instruments in Hamamatsu, Japan. The more localized the rivalry, the more intense. And the more intense, the better.

Another benefit of domestic rivalry is the pressure it creates for constant upgrading of the sources of competitive advantage. The presence of domestic competitors automatically cancels the types of advantage that come from simply being in a particular nation—factor costs, access to or preference in the home market, or costs to foreign competitors who import into the market. Companies are forced to move beyond them, and as a result, gain more sustainable advantages. Moreover, competing domestic rivals will keep each other honest in obtaining government support. Companies are less likely to get

Estimated Number of Japanese Rivals in Selected Industries

Air Conditioners	13
Audio Equipment	25
Automobiles	9
Cameras	15
Car Audio	12
Carbon Fibers	7
Construction Equipment*	15
Copiers	14
Facsimile Machines	10
Large-scale Computers	6
Lift Trucks	8
Machine Tools	112
Microwave Equipment	5
Motorcycles	4
Musical Instruments	4
Personal Computers	16
Semiconductors	34
Sewing Machines	20
Shipbuilding†	33
Steel‡	5
Synthetic Fibers	8
Television Sets	15
Truck and Bus Tires	5
Trucks	11
Typewriters	14
Videocassette Recorders	10

Sources: Field interviews; *Nippon Kogyo Shinbun, Nippon Kogyo Nenkan*, 1987; Yano Research, *Market Share Jitan*, 1987; researchers' estimates.

*The number of companies varied by product area. The smallest number, 10, produced bulldozers. Fifteen companies produced shovel trucks, truck cranes, and asphalt-paving equipment. There were 20 companies in hydraulic excavators, a product area where Japan was particularly strong.

†Six companies had annual production exports in excess of 10,000 tons.

‡Integrated companies.

hooked on the narcotic of government contracts or creeping industry protectionism. Instead, the industry will seek—and benefit from—more constructive forms of government support, such as assistance in opening foreign markets, as well as investments in focused educational institutions or other specialized factors.

Ironically, it is also vigorous domestic competition that ultimately pressures domestic companies to look at global markets and toughens them to succeed in them. Particularly when there are economies of scale, local competitors force each other to look outward to foreign markets to capture greater efficiency and higher profitability. And having been tested by fierce domestic competition, the stronger companies are well equipped to win abroad. If Digital Equipment can hold its own against IBM, Data General, Prime, and Hewlett-Packard, going up against Siemens or Machines Bull does not seem so daunting a prospect.

The Diamond as a System

Each of these four attributes defines a point on the diamond of national advantage; the effect of one point often depends on the state of others. Sophisticated buyers will not translate into advanced products, for example, unless the quality of human resources permits companies to meet buyer needs. Selective disadvantages in factors of production will not motivate innovation unless rivalry is vigorous and company goals support sustained investment. At the broadest level, weaknesses in any one determinant will constrain an industry's potential for advancement and upgrading.

But the points of the diamond are also self-reinforcing: they constitute a system. Two elements, domestic rivalry and geographic concentration, have especially great power to transform the diamond into a system—domestic rivalry because it promotes improvement in all the other determinants and geographic concentration because it elevates and magnifies the interaction of the four separate influences.

The role of domestic rivalry illustrates how the diamond operates as a self-reinforcing system. Vigorous domestic rivalry stimulates the development of unique pools of specialized factors, particularly if the rivals are all located in one city or region: the University of California at Davis has become the world's leading center of wine-making research, working closely with the California wine industry. Active local rivals also upgrade domestic demand in an industry. In furniture and shoes, for example, Italian consumers have learned to expect more and better products because of the rapid pace of new product development that is driven by intense domestic rivalry among hundreds of Italian companies. Domestic rivalry also promotes the formation of related and supporting industries. Japan's world-leading group of semiconductor producers, for instance, has spawned world-leading Japanese semiconductor-equipment manufacturers.

The effects can work in all directions: sometimes world-class suppliers become new entrants in the industry they have been supplying. Or highly sophisticated buyers may themselves enter a supplier industry, particularly when they have relevant skills and view the new industry as strategic. In the case of the Japanese robotics industry, for example, Matsushita and Kawasaki originally designed robots for internal use before beginning to sell robots to others. Today they are strong competitors in the robotics industry. In Sweden, Sandvik moved from specialty steel into rock drills, and SKF moved from specialty steel into ball bearings.

Another effect of the diamond's systemic nature is that nations are rarely home to just one competitive industry; rather, the diamond creates an environment that promotes *clusters* of competitive industries. Competitive industries are not scattered helter-skelter throughout the economy but are usually linked together through vertical (buyer-seller) or horizontal (common customers, technology, channels) relationships. Nor are clusters usually scattered physically; they tend to be concentrated geographically. One competitive industry helps to create another in a mutually reinforcing process. Japan's strength in consumer electronics, for example, drove its success in semiconductors toward the memory chips and integrated circuits these products use. Japanese strength in laptop computers, which contrasts to limited success in other segments, reflects the base of strength in other compact, portable products and leading expertise in liquid-crystal display gained in the calculator and watch industries.

Once a cluster forms, the whole group of industries becomes mutually supporting. Benefits flow forward, backward, and horizontally. Aggressive rivalry in one industry spreads to others in the cluster, through spin-offs, through the exercise of bargaining power, and through diversification by established companies. Entry from other industries within the cluster spurs upgrading by stimulating diversity in R&D approaches and facilitating the introduction of new strategies and skills. Through the conduits of suppliers or customers who have contact with multiple competitors, information flows freely and innovations diffuse rapidly. Interconnections within the cluster, often unanticipated, lead to perceptions of new ways of competing and new opportunities. The cluster becomes a vehicle for maintaining diversity and overcoming the inward focus, inertia, inflexibility, and accommodation among rivals that slows or blocks competitive upgrading and new entry.

The Role of Government

In the continuing debate over the competitiveness of nations, no topic engenders more argument or creates less understanding than the role of the government. Many see government as an essential helper or supporter of industry, employing a host of policies to contribute directly to the competitive performance of strategic or target industries. Others accept the "free market" view that the operation of the economy should be left to the workings of the invisible hand.

Both views are incorrect. Either, followed to its logical outcome, would lead to the permanent erosion of a country's competitive capabilities. On one hand, advocates of government help for industry frequently propose policies that would actually hurt companies in the long run and only create the demand for more helping. On the other hand, advocates of a diminished government presence ignore the legitimate role that government plays in shaping the context and institutional structure surrounding companies and in creating an environment that stimulates companies to gain competitive advantage.

Government's proper role is as a catalyst and challenger; it is to encourage—or even push—companies to raise their aspirations and move to higher levels of competitive performance, even though this process may be inherently unpleasant and difficult. Government cannot create competitive industries; only companies can do that. Government plays a role that is inherently partial, that succeeds only when working in tandem with favorable underlying conditions in the diamond. Still, government's role of transmitting and amplifying the forces of the diamond is a powerful one. Government policies that succeed are those that create an environment in which companies can gain competitive advantage rather than those that involve government directly in the process, except in nations early in the development process. It is an indirect, rather than a direct, role.

Japan's government, at its best, understands this role better than anyone—including the point that nations pass through stages of competitive development and that government's appropriate role shifts as the economy progresses. By stimulating early demand for advanced products, confronting industries with the need to pioneer frontier technology through symbolic cooperative projects, establishing prizes that reward quality, and pursuing other policies that magnify the forces of the diamond, the Japanese government accelerates the pace of innovation. But like government officials anywhere, at their worst Japanese bureaucrats can make the same mistakes: attempting to manage industry structure, protecting the market too long, and yielding to political pressure to insulate inefficient retailers, farmers, distributors, and industrial companies from competition.

It is not hard to understand why so many governments make the same mistakes so often in pursuit of national competitiveness: competitive time for companies and political time for governments are fundamentally at odds. It often takes more than a decade for an industry to create competitive advantage; the process entails the long upgrading of human skills, investing in products and processes, building clusters, and penetrating foreign markets. In the case of the Japanese auto industry, for instance, companies made their first faltering steps toward exporting in the 1950s—yet did not achieve strong international positions until the 1970s.

But in politics, a decade is an eternity. Consequently, most governments favor policies that offer easily perceived short-term benefits, such as subsidies, protection, and arranged mergers—the very policies that retard innovation. Most of the policies that would make a real difference either are too slow and require too much patience for politicians or, even worse, carry with them the sting of short-term pain. Deregulating a protected industry, for example, will lead to bankruptcies sooner and to stronger, more competitive companies only later.

Policies that convey static, short-term cost advantages but that unconsciously undermine innovation and dynamism represent the most common and most profound error in government industrial policy. In a desire to help, it is all too easy for governments to adopt policies such as joint projects to avoid "wasteful" R&D that undermine dynamism and competition. Yet even a 10% cost saving through economies of scale is easily nullified through rapid product and process improvement and the pursuit of volume in global markets—something that such policies undermine.

There are some simple, basic principles that governments should embrace to play the proper supportive role for national competitiveness: encourage change, promote domestic rivalry, stimulate innovation. Some of the specific policy approaches to guide nations seeking to gain competitive advantage include the following.

Focus on specialized factor creation. Government has critical responsibilities for fundamentals like the primary and secondary education systems, basic national infrastructure, and research in areas of broad national concern such as health care. Yet these kinds of generalized efforts at factor creation rarely produce competitive advantage. Rather, the factors that translate into competitive advantage are advanced, specialized, and tied to specific industries or industry groups. Mechanisms such as specialized apprenticeship programs, research efforts in universities connected with an industry, trade association activities, and, most important, the private investments of companies ultimately create the factors that will yield competitive advantage.

Avoid intervening in factor and currency markets. By intervening in factor and currency markets, governments hope to create lower factor costs or a favorable exchange rate that will help companies compete more effectively in international markets. Evidence from around the world indicates that these policies—such as the Reagan administration's dollar devaluation—are often counterproductive. They work against the upgrading of industry and the search for more sustainable competitive advantage.

The contrasting case of Japan is particularly instructive, although both Germany and Switzerland have had similar experiences. Over the past 20 years, the Japanese have been rocked by the sudden Nixon currency devaluation shock, two oil shocks, and, most recently, the yen shock—all of which forced Japanese companies to upgrade their competitive advantages. The point is not that government should pursue policies that intentionally drive up factor costs or the exchange rate. Rather, when market forces create rising factor costs or a higher exchange rate, government should resist the temptation to push them back down.

Enforce strict product, safety, and environmental standards. Strict government regulations can promote competitive advantage by stimulating and upgrading domestic demand. Stringent standards for product performance, product safety, and environmental impact pressure companies to improve quality, upgrade technology, and provide features that respond to consumer and social demands. Easing standards, however tempting, is counterproductive.

When tough regulations anticipate standards that will spread internationally, they give a nation's companies a head start in developing products and services that will be valuable elsewhere. Sweden's strict standards for environmental protection have promoted competitive advantage in many industries. Atlas Copco, for example, produces quiet compressors that can be used in dense urban areas with minimal disruption to residents. Strict standards, however, must be combined with a rapid and streamlined regulatory process that does not absorb resources and cause delays.

Sharply limit direct cooperation among industry rivals. The most pervasive global policy fad in the competitiveness arena today is the call for more cooperative research and industry consortia. Operating on the belief that independent research by rivals is wasteful and duplicative, that collaborative efforts achieve economies of scale, and that individual companies are likely to underinvest in R&D because they cannot reap all the benefits, governments have embraced the idea of more direct cooperation. In the United States, antitrust laws have been modified to allow more cooperative R&D; in Europe, megaprojects such as ESPRIT, an information-technology project, bring together companies from several countries. Lurking behind much of this thinking is the fascination of Western governments with—and fundamental misunderstanding of—the countless cooperative research projects sponsored by the Ministry of International Trade and Industry (MITI), projects that appear to have contributed to Japan's competitive rise.

But a closer look at Japanese cooperative projects suggests a different story. Japanese companies participate in MITI projects to maintain good relations with MITI, to preserve their corporate images, and to hedge the risk that competitors will gain from the project—largely defensive reasons. Companies rarely contribute their best scientists and engineers to cooperative projects and usually spend much more on their own private research in the same field. Typically, the government makes only a modest financial contribution to the project.

The real value of Japanese cooperative research is to signal the importance of emerging technical areas and to stimulate proprietary company research. Cooperative projects prompt companies to explore new fields and boost internal R&D spending because companies know that their domestic rivals are investigating them.

Under certain limited conditions, cooperative research can prove beneficial. Projects should be in areas of basic product and process research, not in subjects closely connected to a company's proprietary sources of advantage. They should constitute only a modest portion of a company's overall research program in any given field. Cooperative research should be only indirect, channeled through independent organizations to which most industry participants have access. Organizational structures, like university labs and centers of excellence, reduce management problems and minimize the risk to rivalry. Finally, the most useful cooperative projects often involve fields that touch a number of industries and that require substantial R&D investments.

Promote goals that lead to sustained investment. Government has a vital role in shaping the goals of investors, managers, and employees through policies in various areas. The manner in which capital markets are regulated, for example, shapes the incentives of investors and, in turn, the behavior of companies. Government should aim to encourage sustained investment in human skills, in innovation, and in physical assets. Perhaps the single most powerful tool for raising the rate of sustained investment in industry is a tax incentive for long-term (five years or more) capital gains restricted to new investment in corporate equity. Long-term capital gains incentives should also be applied to pension funds and other currently untaxed investors, who now have few reasons not to engage in rapid trading.

Deregulate competition. Regulation of competition through such policies as maintaining a state monopoly, controlling entry into an industry, or fixing prices has two strong negative consequences: it stifles rivalry and innovation as companies become preoccupied with dealing with regulators and protecting what they already have; and it makes the industry a less dynamic and less desirable buyer or supplier. Deregulation and privatization on their own, however, will not succeed without vigorous domestic rivalry—and that requires, as a corollary, a strong and consistent antitrust policy.

Enforce strong domestic antitrust policies. A strong antitrust policy—especially for horizontal mergers, alliances, and collusive behavior—is fundamental to innovation. While it is fashionable today to call for mergers and alliances in the name of globalization and the creation of national champions, these often undermine the creation of competitive advantage. Real national competitiveness requires governments to disallow mergers, acquisitions, and alliances that involve industry leaders. Furthermore, the same standards for mergers and alliances should apply to both domestic and foreign companies. Finally, government policy should favor internal entry, both domestic and international, over acquisition. Companies should, however, be allowed to acquire small companies in related industries when the move promotes the transfer of skills that could ultimately create competitive advantage.

Reject managed trade. Managed trade represents a growing and dangerous tendency for dealing with the fallout of national competitiveness. Orderly marketing agreements, voluntary restraint agreements, or other devices that set quantitative targets to divide up markets are dangerous, ineffective, and often enormously costly to consumers. Rather than promoting innovation in a nation's industries, managed trade guarantees a market for inefficient companies.

Government trade policy should pursue open market access in every foreign nation. To be effective, trade policy should not be a passive instrument; it cannot respond only to complaints or work only for those industries that can muster enough political clout; it should not require a long history of injury or serve only distressed industries. Trade policy should seek to open markets wherever a nation has competitive advantage and should actively address emerging industries and incipient problems.

Where government finds a trade barrier in another nation, it should concentrate its remedies on dismantling barriers, not on regulating imports or exports. In the case of Japan, for example, pressure to accelerate the already rapid growth of manufactured imports is a more effective approach than a shift to managed trade. Compensatory tariffs that punish companies for unfair trade practices are better than market quotas. Other increasingly important tools to open markets are restrictions that prevent companies in offending nations from investing in acquisitions or production facilities in the host country—thereby blocking the unfair country's companies from using their advantage to establish a new beachhead that is immune from sanctions.

Any of these remedies, however, can backfire. It is virtually impossible to craft remedies to unfair trade practices that avoid both reducing incentives for domestic companies to innovate and export and harming domestic buyers. The aim of remedies should be adjustments that allow the remedy to disappear.

The Company Agenda

Ultimately, only companies themselves can achieve and sustain competitive advantage. To do so, they must act on the fundamentals described above. In particular, they must recognize the central role of innovation—and the uncomfortable truth that innovation grows out of pressure and challenge. It takes leadership to create a dynamic, challenging environment. And it takes leadership to recognize the all-too-easy escape routes that appear to offer a path to competitive advantage, but are actually short-cuts to failure. For example, it is tempting to rely on cooperative research and development projects to lower the cost and risk of research. But they can divert company attention and resources from proprietary research efforts and will all but eliminate the prospects for real innovation.

Competitive advantage arises from leadership that harnesses and amplifies the forces in the diamond to promote innovation and upgrading. Here are just a few of the kinds of company policies that will support that effort:

Create pressures for innovation. A company should seek out pressure and challenge, not avoid them. Part of strategy is to take advantage of the home nation to create the impetus for innovation. To do that, companies can sell to the most sophisticated and demanding buyers and channels; seek out those buyers with the most difficult needs; establish norms that exceed the toughest regulatory hurdles or product standards; source from the most advanced suppliers; treat employees as permanent in order to stimulate upgrading of skills and productivity.

Seek out the most capable competitors as motivators. To motivate organizational change, capable competitors and respected rivals can be a common enemy. The best managers always run a little scared; they respect and study competitors. To stay dynamic, companies must make meeting challenge a part of the organization's norms. For example, lobbying against strict product standards signals the organization that company leadership has diminished aspirations. Companies that value stability, obedient customers, dependent suppliers, and sleepy competitors are inviting inertia and, ultimately, failure.

Establish early-warning systems. Early-warning signals translate into early-mover advantages. Companies can take actions that help them see the signals of change and act on them, thereby getting a jump on the competition. For example, they can find and serve those buyers with the most anticipatory needs; investigate all emerging new buyers or channels; find places whose regulations foreshadow emerging regulations elsewhere; bring some outsiders into the management team; maintain ongoing relationships with research centers and sources of talented people.

Improve the national diamond. Companies have a vital stake in making their home environment a better platform for international success. Part of a company's responsibility is to play an active role in forming clusters and to work with its home-nation buyers, suppliers, and channels to help them upgrade and extend their own competitive advantages. To upgrade home demand, for example, Japanese musical instrument manufacturers, led by Yamaha, Kawai, and Suzuki, have established music schools. Similarly, companies can stimulate and support local suppliers of important specialized inputs—including encouraging them to compete globally. The health and strength of the national cluster will only enhance the company's own rate of innovation and upgrading.

In nearly every successful competitive industry, leading companies also take explicit steps to create specialized factors like human resources, scientific knowledge, or infrastructure. In industries like wool cloth, ceramic tiles, and lighting equipment, Italian industry associations invest in market information, process technology, and common infrastructure. Companies can also speed innovation by putting their headquarters and other key operations where there are concentrations of sophisticated buyers, important suppliers, or specialized factor-creating mechanisms, such as universities or laboratories.

Welcome domestic rivalry. To compete globally, a company needs capable domestic rivals and vigorous domestic rivalry. Especially in the United States and Europe today, managers are wont to complain about excessive competition and to argue for mergers and acquisitions that will produce hoped-for economies of scale and critical mass. The complaint is only natural—but the argument is plain wrong. Vigorous domestic rivalry creates sustainable competitive advantage. Moreover, it is better to grow internationally than to dominate the domestic market. If a company wants an acquisition, a foreign one that can speed globalization and supplement home-based advantages or offset home-based disadvantages is usually far better than merging with leading domestic competitors.

Globalize to tap selective advantages in other nations. In search of "global" strategies, many companies today abandon their home diamond. To be sure, adopting a global perspective is important to creating competitive advantage. But relying on foreign activities that supplant domestic capabilities is always a second-best solution. Innovating to offset local factor disadvantages is better than outsourcing; developing domestic suppliers and buyers is better than relying solely on foreign ones. Unless the critical underpinnings of competitiveness are present at home, companies will not sustain competitive advantage in the long run. The aim should be to upgrade home-base capabilities so that foreign activities are selective and supplemental only to over-all competitive advantage.

The correct approach to globalization is to tap selectively into sources of advantage in other nations' diamonds. For example, identifying sophisticated buyers in other countries helps companies understand different needs and creates pressures that will stimulate a faster rate of innovation. No matter how favorable the home diamond, moreover, important research is going on in other nations. To take advantage of foreign research, companies must station high-quality people in overseas bases and mount a credible level of scientific effort. To get anything back from foreign research ventures, companies must also allow access to their own ideas—recognizing that competitive advantage comes from continuous improvement, not from protecting today's secrets.

Use alliances only selectively. Alliances with foreign companies have become another managerial fad and cure-all: they represent a tempting solution to the problem of a company wanting the advantages of foreign enterprises or hedging against risk, without giving up independence. In reality, however, while alliances can achieve selective benefits, they always exact significant costs: they involve coordinating two separate operations, reconciling goals with an independent entity, creating a competitor, and giving up profits. These costs ultimately make most alliances short-term transitional devices, rather than stable, long-term relationships.

Most important, alliances as a broad-based strategy will only ensure a company's mediocrity, not its international leadership. No company can rely on another outside, independent company for skills and assets that are central to its competitive advantage. Alliances are best used as a selective tool, employed on a temporary basis or involving noncore activities.

Locate the home base to support competitive advantage. Among the most important decisions for multinational companies is the nation in which to locate the home base for each distinct business. A company can have different home bases for distinct businesses or segments. Ultimately, competitive advantage is created at home: it is where strategy is set, the core product and process technology is created, and a critical mass of production takes place. The circumstances in the home nation must support innovation; otherwise the company has no choice but to move its home base to a country that stimulates innovation and that provides the best environment for global competitiveness. There are no half-measures: the management team must move as well.

The Role of Leadership

Too many companies and top managers misperceive the nature of competition and the task before them by focusing on improving financial performance, soliciting government assistance, seeking stability, and reducing risk through alliances and mergers.

Today's competitive realities demand leadership. Leaders believe in change; they energize their organizations to innovate continuously; they recognize the importance of their home country as integral to their competitive success and work to upgrade it. Most important, leaders recognize the need for pressure and challenge. Because they are willing to encourage appropriate—and painful—government policies and regulations, they often earn the title "statesmen," although few see themselves that way. They are prepared to sacrifice the easy life for difficulty and, ultimately, sustained competitive advantage. That must be the goal, for both nations and companies: not just surviving, but achieving international competitiveness.

And not just once, but continuously.

WHARTON REPROGRAPHICS

Strategic Management Journal, Vol. 12, 33–47 (1991)

COUNTRY CAPABILITIES AND THE PERMEABILITY OF BORDERS

BRUCE KOGUT
The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania, U.S.A.

The history of international competition shows long cycles in country leadership. This paper argues that these cycles can be explained by differences in country capabilities, in terms of both technologies and organizing principles, which diffuse more slowly across than within national and regional borders. The country that innovates in best practices eventually loses leadership, but not before increasing its wealth and claims on foreign assets and investments.

International competition has frequently been analyzed in terms of the relative costs and abundance of the factors of production, and by the conditions of rivalry in national and world markets. Until recently, little attention was paid directly to the circumstances that explain why a country should be favored across a wide variety of industries.[1] Current treatments tend to emphasize the conditions of industry structure over the broad accumulation of country capabilities as the critical point of inquiry.

This article proposes that the notion of country competitiveness rests on a disarmingly simple question: are the borders of a country less permeable than the borders of the firm? On inspection, the kernel of the problem boils down to a fundamental issue of why superior practices are not more readily adopted across borders of countries. It is this question that lies at the heart of the important debate on the capabilities of nations and the international competitiveness of firms.

The discussion below is driven by two stylized facts. The first is that countries differ in their prevailing technological and organizational capabilities. The second is that organizational capabilities diffuse more slowly than technologies. Together, these two facts imply that competitive differences among countries should persist for long periods of time due primarily to the slow adoption of organizing principles. But this implication still skirts the more troubling question of why organizational capabilities should diffuse more rapidly within a country than among countries.

The analysis of this question provides a window by which to understand long-term patterns in country competitiveness. We propose that countries differ in their underlying 'organizing principles' of work, and that these principles develop within the confines of a trajectory. These principles diffuse within an inter-industry network of firms, which, though differing in their products and markets, share common heuristics as applied to how their economic activities are organized. In this sense the diffusion of organizing principles embodies the know-how defining a country's capabilities.

The speed of diffusion of new methods raises the question of whether these national methods can be identified, understood, and assimilated across borders. Organizational innovations, such as a new computer system, may be fairly

---

[1] See Dunning (1990), Kogut (1987) and Porter (1990) as recent exceptions, and the substantial literature in comparative management and the sociology of work more generally.

0143–2095/91/050033–15$07.50
© 1991 by John Wiley & Sons, Ltd.

transparent in terms of the technical information, but their implementation involves complex interdependence with existing organizational structures and operation. It is harder to transfer, or imitate knowledge, when their causal properties are poorly understood and articulated in terms of their implications for the prevailing organizing principles of a firm and their relations with other institutions. Ultimately, the analysis of the permeability of firm and national borders hinges on the institutional characteristics of the organizing principles of a country.

Based on the notions of organizing principles developing within a trajectory, we reach the following conclusions:

1. Trade and foreign direct investment patterns among developed countries reflect the sectors favored relatively by a country's organizing and technological strengths, and these patterns promote further expansion and investment in these capabilities.
2. The driving explanation for sustained absolute country advantages is to be found in the persisting variation of organizational and institutional capabilities among countries.
3. Horizontal foreign direct investment is the extension of organizing capabilities across borders.
4. Vertical foreign direct investment is the sourcing of cheaper factors of production, as well as the tapping of the resident technological and organizing principles of foreign locations.
5. Long-term cycles in country leadership are generated by the arrival of new methods of organization that are first limited to a country and then slowly diffuse internationally, but, ultimately, the leading country has gained in wealth, inclusive of foreign assets.

These points are developed below by drawing on disparate, yet overlapping literatures, with the principal sources being Lippman and Rumelt (1982), Nelson and Winter (1982), Piore and Sabel (1984), Stiglitz (1987), and Winter (1987). The first sections propose that countries are characterized by the accumulation of capabilities and that competitive differences among countries persist for long periods of time. The next section broadens this perspective to include the organizing principles that prevail in a country. The concept of neighboring technologies is examined in the fourth section in terms of the institutionalization of technology and organizing principles. The fifth section turns to explaining the slow convergence to best practice in terms of four factors: technological opportunity, selection pressures, identifiability and institutional lock-in. The final section explores why the slow diffusion of best practices and differences in learning capabilities generate competitive cycles of leading countries.

LONG-TERM TRENDS IN COUNTRY COMPETITION

Concern over the relative decline of the United States during the 1970s and 1980s has focused attention on recent history as period of inquiry. There has been unquestionably a perplexing productivity problem in the United States since the oil shocks, though its severity is not greatly different from that found in other developed countries, including Japan. And there is certainly little question that the U.S. dominance has diminished greatly since the early post-war period.

However, as the seminal study by Maddison (1982) shows, the most impressive years for the U.S. were those of the late 1800s and early 1900s. The trends point very clearly to the emergence of the U.S. in the late 1800s, with the U.K. falling steadily behind. Equally impressive is the rise of Germany and Japan at the turn of this century. The effects of wartime losses are apparent, but what is remarkable is the resumption of these trends after both world wars for Germany, and after World War II for Japan.

Because productivity has been the center of considerable controversy, it is especially interesting to note that the U.S. has never had an outstanding performance in labor productivity in any one period. While its persistent cumulative growth over a century is striking, there is little doubt that the productivity lead of the United States had been eliminated by the 1980s. This loss of leadership contrasts strongly with the dominant position held by the U.S. earlier in the century. Sectoral studies indicate, moreover, that this productivity advantage was diffused across many industries. In particular, in a study that received substantial attention in the U.K. at the time of its publication, Rostas (1948) found that the U.S. had better productivity than the U.K.

in every industry he examined; even as recently as 1972 the U.S. led Japan in 54 of 60 industries (Maddison, 1982: 102–103).

These findings are important reminders of the distinction of relative and absolute advantage. Since Ricardo, it is well established that the sectoral composition of trade among countries is determined by relative advantage. A country will specialize in those industries in which it is relatively best at doing. Investment allocations are influenced by home and foreign demand; the development of relatively unfavored industries over time is, consequently, hindered. The observation that a country has a comparative advantage in exporting a particular product is not an indication of any absolute country advantage, as prices (perhaps via exchange rate movements) eventually equilibrate trade flows among countries. But the equilibrium will reflect absolute advantage by differences in the wealth of countries. Ricardo's lesson of relative advantage is a statement about the sectoral composition of trade, not an explanation of an absolute country advantage; to the contrary, it leaves unanswered why a country may be absolutely more productive in all sectors.

TECHNOLOGICAL CAPABILITIES

These long-term trends pose a fundamental quandry for theories of international trade and investment that focus upon sectoral composition. Most explanations for export and foreign direct investment flows have tended to focus on the industry and firm level of analysis. Home market rivalry and a firm's possession of 'intangible' assets, such as brand labels or product innovations, drive the international competitiveness of a country's industries.

The most consistent empirical finding has been the correlation between R&D expenditures and the international competitiveness of industries. In a recent review of empirical work on the determinants of exports, Deardorff (1984) concludes that R&D is the primary determinant of exports. Similarly, almost all empirical studies on foreign direct investment have found R&D expenditures to be significantly correlated with outward flows.[2]

But this correlation of R&D expenditures with exports or foreign direct investment has an important implication that is frequently missed. R&D is as much an outcome of profits, and growth, as a determinant; distilling the causal path has proven empirically difficult (Griliches, 1986). Countries develop and reinvest in those industries in which they are relatively favored. What starts out as relative advantage determining trade patterns itself over time influences the investment in new advantages. Comparative advantage and the capabilities of countries are dynamically altered by new investment.

In this regard, it is an important observation that the pattern of technological specialization among countries appears to be stable over long periods of time. Pavitt and Soete (1982) found distinct differences in the sectoral distribution of patents and export shares among countries. In extending the work on technological accumulation to country competitiveness, Pavitt (1987) found that several measures of a country's technical specialization are stable over several decades. In testing for this stability, Cantwell (1989) estimated that country shares of patents among industry sectors are significantly correlated during the post-World War II period and, to a lesser extent, during the century.

These findings are especially significant in light of recent arguments regarding the cumulativeness of technology at the firm level.[3] They pose, therefore, the question of whether technology accumulates in a country in much the same way that it appears to accumulate in firms. What is interesting in considering this question is that the answer points to the importance of the institutional context and the organizing principles of a country.

Consider first, the explanation for why technology accumulates at the firm level and generates firm-specific capabilities. The development of firm capabilities is often viewed as indicating that technological advancement occurs by incremental additions to the current stock of knowledge. In this regard a more ambitious claim is that technology is not only incremental, but also follows a 'natural trajectory' (Nelson and Winter, 1977). Firms search, in the Cyert and March (1963) sense, in the neighborhood of current practice for solutions to problems and opportunities as they arrive and are identified. Innovations are made in response to market demand, including the demand to substitute scarce or costly factors.[4] Moreover, their viability is also determined by market acceptance. In this sense demand plays an instrumental role in the production and success of innovative activity.

---

[2] See Caves (1982).

[3] See, for example, Nelson and Winter (1977, 1982) and Dosi (1982).

Once, however, a set of technologies are selected, they are constrained to evolve within the limits of physical laws, but are also successively applied to new problems in related markets and industries. Thus, the notion of a technological trajectory is closely coupled to the concept of technological opportunities, as developed by Scherer (1965). These opportunities are driven forward by the expansion of basic sciences and by the applicability to new industries. A new trajectory is the routinized application of novel solution heuristics that expresses the current frontier knowledge.

These observations on the cumulativeness of technology embed implicitly the argument for why firms within a country develop similar technological capabilities. In part the commonality stems from the basic assumption that firms respond to market needs and that the market selects products (and potentially firms) on the basis of their competitive cost/performance qualities. The effect of facing common market stimuli and demand-driven forces of selection can alone explain why a population of firms within a given environment tend to offer similar products.[5] Across-country variation in demand characteristics, by this argument, generates a parallel variation in product types.

But this argument of market selection is not sufficient to explain why these home market patterns should extend to international trade. Nor does it address the persistence in country technological patterns. It is the cumulative process by which technological capabilities are built that explains the link between trade patterns and the home market. The cumulative capabilities of firms, developed in response to their home markets, provide the competitive basis for expansion overseas, yet at the same time, limit the feasible range of products. Foreign trade and investment, as Burenstam Linder (1961) seminally observed, are but the extension of the home market across borders.

Of course, firms whose origins occur under conditions of the internationalization of markets may develop a wider set of capabilities than those corresponding to the needs of any one national market. But this logical possibility only underscores the argument, that is, that the early entrepreneurial history of a firm tends to 'lock in' its subsequent development.[6] That most firms have originated in environments that are largely national is an empirical claim, but one that has tended to be historically true (Porter, 1990).

Trade and investment flows consist, then, of the domestic products that correspond to the international market. Yet, as noted earlier, there is a dynamic element, as the expansion overseas draws investments and new resources into these favored industries. Thus, the well-known property of international trade tending towards specialization also has the implication of fostering the specialization of a country's capabilities.

The discussion so far has stylized the problem as the accumulation of capabilities lodged within firms. This view is too restrictive, for capabilities are also developed by the wider set of institutions in a country. For example, the strength of the German universities in chemistry was noted in the late 1800s, especially as the flow of German students to British universities began to ebb (Freeman, 1982). The link between science-based centers and economic organizations is one of the institutional aspects constituting national systems of innovation and technology (Freeman, 1987).[7]

In this wider perspective, then, the investments made by public institutions are part of the process by which capabilities are routinized within a country. The characteristics of these systems differ across countries in terms of human capital formation, the level of entrepreneurial activity, or the rigorousness of competition. But also since the technological opportunities of different sciences will vary, the long-term performance of countries can vary to the extent that these scientific resources are national and do not spill over borders.

---

[4] That innovation is stimulated by scarcity has a long history of controversy, including in the area of international economics. See David (1975) for a general discussion, Winter (1981) for a formal treatment, and Kogut (1987) for a review of the international applications.

[5] This argument does not rule out learning among firms. Selection may operate at the product level, with firms learning from each other's success.

[6] See Arthur (1989) and Stiglitz (1987). Kogut (1987) discusses a similar idea labeled 'intertemporal dependence'.

[7] See also Malerba (1990) for a description of the Italian system of innovations and Herrigel (1992) for an analysis of the institutional resources in the German machine tool industry.

COUNTRY CAPABILITIES AND ORGANIZING PRINCIPLES

There is, of course, no reason to limit this notion of a trajectory spreading within an economy to a narrow definition of technology. Changes in the organization of work, either at the factory, corporate, or industry level, can also follow a natural trajectory. In fact, the two examples given by Nelson and Winter (1977) of a trajectory are essentially organizational: that is, the exploitation of latent scale economies and increasing mechanization of operations that have been done by hand.

It is, of course, difficult to establish the numerical data, as can be established for R&D, indicating country specialization in organizing principles. These principles may be of a few types, such as the craft and mass production distinction made by Piore and Sabel (1984) and Coriat (1979). Or they may be more variegated in terms of structural characteristics, as suggested in several bi-national studies,[8] or of managerial ideologies (Bendix, 1956).[9] But that such differences exist can scarcely be doubted.

A more provocative claim is that organizational innovations also develop along a trajectory. Several historical accounts suggest such a dynamic in the slow diffusion of organizational practices within a country. Consider three examples of the diffusion of organizing principles from the industrial histories of the United Kingdom, the United States and Japan (their international diffusions are traced later). The advancement of the United Kingdom was due to the expansion of organized industrial production, as Adam Smith so clearly explained in his example of the division of labor in the fabrication of pins. The application of water, and later steam, to drive capital equipment within an organized factory made its early and most striking appearance in the British textile industry before diffusing widely to other industries (Deane, 1975). That the division of labor within the factory was accompanied by the introduction of new capital equipment should not obscure the significance of the organizing innovations.

The U.S. developments that led to a system of mass production are especially well documented.[10] Smith's (1977) study of interchangeability of parts in the U.S. armories showed a 30-year lag before the techniques were transferred successfully among only two armories. The development of the American System of Manufacturing took more than 100 years to evolve the concept of standardization and mass production (Hounshell, 1984). Interestingly, the system evolved from the application of standard gauges in weapons armories used in craft shops to the application of standardization in the very production process itself, as best exemplified in Ford's assembly line. Of great importance, the pattern of diffusion shows the important role played by a few key industries in diffusing these techniques within the national industrial and service network. Similarly, the principles of scientific management–as a further expression of the standardization of production—spread from the machine shops in the Philadelphia area to tire manufacturing to wholesale/retail operations (Nelson, 1980). In the 1920s and 1930s it was actively applied to the office place, as especially expressed in the creation of secretarial pools (Davies, 1982).

A more recent example is derived from Japan, as best characterized by Toyota's use, in Krafcik's (1988), terminology, of lean flexible production, though the historical roots may be much earlier (Dohse, Juergens and Melsch, 1985). The historical accounts indicate that the principles of just-in-time and pull systems (kanban) of manufacturing were innovations taking place at Toyota in the late 1950s before spreading to suppliers and eventually other firms (Cusumano, 1985; Fruen,

---

[8] See, for example, Dore (1973), Lincoln, Hanada and McBride (1986), and Maurice, Sellier and Silvestre (1984).

[9] The genesis of these country differences is an important topic, but peripheral to our argument. The sociological foundation to such an inquiry is best laid out in Stinchcombe's (1965) discussion of the imprinting of the social environment on the structure of organizations. For a discussion of imprinting and country variations, see Kogut (1987).

[10] See the more extensive account in Kogut and Parkinson (1992).

1989). The full documentation is far from complete, but it would not be surprising that these techniques have applications to non-manufacturing sectors as well.

The above historical account has a clear implication, that is, that major innovations of organizing principles occurred in different countries at different moments in history. Certainly, parallel developments have taken place in many countries, as, for example, with the simultaneous application of engineering concepts to the organization of factory work in western Europe and the U.S. (Maier, 1970). Yet it was in the U.S. that scientific management first was widely disseminated, as clearly recognized by Europeans at that time (Devinat, 1927).

Of course, a country can compete on world markets with inferior organizing capabilities, but cheaper factor costs, as the doctrine of relative advantage has long demonstrated. But to return to the puzzle posed by the long-term trends given by Maddison, the dominance of a country's world position requires an explanation consistent with absolute advantage. Certainly, the solution may lie in high rates of capital accumulation, natural resources, or labor abundance, but the econometric evidence suggests that this answer leaves much to be explained (Nelson, 1981).

The explanation outlined above is that the leadership of a country is not driven alone by technological investments, but also by the efficiency of the dominant organizing principles.[11] There is little evidence, for example, that the U.S. economy was performing better in the production of technologies (as measured by patents) than the German or the U.K. (Cantwell, 1989). In the view of Boyer (1984), the slowdown of European productivity growth and the superior performance of the American economy just prior to World War I were attributable to the adoption of new methods of management in the U.S., breaking the bottleneck in the organization of factory work that was plaguing both continents.

A cursory analysis of world competition in automobiles—one of the most important manufacturing industries in international trade and investment in this century—is telling. The failure of U.S. bicycle manufacturers to consolidate factories into mass production and maintain their export dominance is in counterpoint to auto companies, even though the development of automobiles and the breakthrough technologies were pioneered in Europe and, especially, France (Laux, 1976). Ford, in particular, led the way, with the leading European market share supplied by a single plant in the U.K. by World War I. The response of the French was a constant effort to adopt American methods (Fridenson, 1987), an effort which largely failed in the U.K. (Lewchuck, 1987). Similarly, Ford and General Motors, both of which exported knock-down vehicles from the U.S., had far away dominant market share in Japan through the 1930s (Mason, 1990).

On reflection, it is hardly accidental that no European producer relying on American-derived techniques of mass production has achieved significant penetration in the U.S., with a brief exception of Volkswagen before German wages climbed to parity. The success of the European auto companies in the U.S. has largely been achieved through the export of higher-end vehicles, manufactured on the basis of craft modes of production, as described by Piore and Sabel (1984). Nor did Japanese auto companies achieve dramatic success in the U.S. in the 1950s and 1960s, despite considerable efforts to do so and much lower domestic wages. Japanese penetration in American markets occurred only subsequent to the further evolution and application of the principles of 'Toyotism' (Cusumano, 1985; Dohse *et al.*, 1985).

The auto history highlights an important point. Foreign direct investment is the extension of the organizing principles of domestic firms into foreign markets. As has been often noted, multinational corporations carry their home country practices overseas.[12] This point was not lost on the European observers such as Servan-Schreiber (1969), who noted that the U.S. success in Europe was largely based on 'the new methods of organization' embodied in its domestic multinational corporations. With the penetration of Japanese trade and multinational investments in the U.S. and Europe, observations and empirical evidence on the importance of the related organizing practices were prevalent (Dunning, 1986; Ouchi, 1981). Parenthetically, it is important to note that this unidirectional extension of country capabilities via exports and foreign direct investment becomes attenuated over time by the activities of multinational corporations. Dominant theories of foreign direct investment emphasize the 'push' of home industry and firm factors on the horizontal expansion of a firm across borders. The above argument proposes that country capabilities can also 'pull' the location of foreign subsidiaries for the sourcing of resident advantages in technologies and organizing practices. By implication, the puzzle of the tendency of foreign direct investment in manufacturing to flow towards high labor countries can be resolved as the product of the pulling of investment to the locational centers of the leading technologies and best practices.[13] But, as consistent with our earlier argument, we would expect that the sourcing—and especially transfer—of best practice is more fraught with difficulty than the acquisition of technologies.

---

[11] Chandler (1990) is clearly in agreement in the primacy of organizational capability for international competitiveness, but he differs in citing scale and scope as universally instrumental rather than treating organizing principles as differing 'modally' among countries, as best identified with the work of Piore and Sabel (1984).

[12] See, for example, Dunning (1958, 1986) and McInnes (1971).

In summary, the argument that countries differ in their organizational, as well as technological, capabilities, and that these capabilities tend to develop and diffuse within a trajectory, provides a perspective for understanding long-term trends in international competitiveness. International trade and investment are driven by the incorporation of frontier technologies in capital goods and products, and by the international diffusion of best practice in the organizing principles of multinational corporations.

INSTITUTIONAL BASIS OF COUNTRY CAPABILITIES

It would be misleading, however, to suggest that the cumulative nature of technologies and organizing practices prevents imitation. To the contrary, the evidence points to a surprisingly rapid imitation of incremental product and process technologies.[14] While these studies concerned product and process innovations, imitation lags appear to be longer for organizational innovations, as indicated by several empirical studies (Armour and Teece, 1978; Daft, 1978; Damanpour and Evan, 1984).

Whereas there is ample empirical evidence that imitation times and costs vary by technology, industry, and even country, there is no consensus on the explanation. Certainly, the number of competitors and their resources are a portion of the explanation, but this industry feature then begs the question why the number of imitators varies across industries. In part, the answer is to be found in what can be called the 'conditions of appropriability', that is, the legal and institutional factors which determine the profitability to the innovation which can be recovered by the innovator. As these conditions vary by country (e.g. patent policy or industry structure), the potential for appropriability also varies across nations (Kogut, 1987; Teece, 1987).

Another explanation appeals to history dependence and lock-in to current practice, as described earlier. Atkinson and Stiglitz (1969) observed that learning externalities are 'localized', that is, that improvements in a technique only benefit neighboring technologies; consequently, firms tend to make alterations in their practices in closely related technologies. In the view of Lippman and Rumelt (1982), this lock-in results from the uncertainty of imitators to match the efficiency of incumbents. As Rumelt (1984) subsequently noted, their view has the implication that firms should expand in areas related to their competence. In the general model of Arthur (1989), prior experience and joint economies among users reduce the costs of expanding into related technologies within the same trajectory, inducing an inertia against adopting an initially more costly but eventually more efficient alternative. Stiglitz (1987) has argued that the condition of localized learning, plus the costs of switching technologies, lock a country into a long-term development path. Being good at using the wrong technique can lock a country into a sub-optimal technology.

---

[13] See Swedenborg (1979) and Yamawaki (1990) on the pull of manufacturing to high wage countries; Cantwell (1989) on the relationship of the pull of direct investment to countries leading in the related technologies; and Kogut and Chang (1991) on the use of joint ventures by Japanese firms to enter industries where the U.S. invests proportionally more than Japan in R&D expenditures.

[14] See, for example, Levin, Klevorick and Winter (1987), Mansfield (1985), and Tilton (1971).

Though appealing, these formulations beg the question, in turn, what determines the distance between technologies or, equally so, the degree of the uncertainty of imitation? Ultimately, the analysis must turn to exploring this fundamental assumption. One candidate factor is the nature of the technology; that is, the degree to which it is tacit, codifiable, or part of a system (Rogers, 1971; Winter, 1987). On examination these dimensions are intrinsic properties not of the technology, but of the social knowledge of the firm and its nesting within the wider societal institutions (Kogut and Zander, 1990). Codifiability is relevant only to the extent that the parties know the codes and can communicate.

The issue extends, however, beyond the observation that the comprehensibility of a technology depends on a user's previous experience, educational level, or cultural background. The knowledge of a firm is expressed organizationally and institutionally. Its capabilities consist of its organizing abilities to transform the expertise of individuals and groups into saleable products. These capabilities, as informed by its proprietary data, reside in the routinized expectations built on a cumulative and shared experience. Learning is localized because knowledge is institutionally structured in on-going and enduring relationships. To adopt new practices requires a change in these relationships.

Learning is more difficult when the underlying knowledge is recipe- (i.e. know-how) based as opposed to information. The adoption of new technologies may require training and education of individuals in order to absorb *what* the relevant information is. But the adoption of new ways of doing things is complicated because it entails coordination of *how* groups behave and interact.

Of importance to understanding country capabilities is the recognition that some institutionalized relationships span firms. The capabilities of a firm reside not only in its know-how to transform the expertise of its employees into outputs, but also in the strengths of its institutional relationships with customers, suppliers, or sources of new technologies, e.g. universities. The focus upon domestic competition obscures the significance of these institutional linkages because they are often public goods to national companies. But international competition puts into relief the importance of these wider national capabilities and their contribution to the success of individual firms. In this important sense, then, the capabilities of a firm are nested in the wider institutional capabilities of a country.

INTRA-COUNTRY VERSUS INTER-COUNTRY DIFFUSION

The above discussion of the localness of knowledge as a consequence of the stable structure of relationships within a nation provides the basis for why country borders are less permeable than firm borders. The explanation for the difficulty of deciphering and diffusing organizing principles across national borders can be analyzed in terms of four factors: (1) technological opportunities, (2) selection forces, (3) identifiability, and (4) institutional lock-in. By our earlier discussion these factors define the institutional capability of a country, the process by which new techniques are selected and identified, and the degree of inertia of the current knowledge.

Technological opportunities

Technological opportunities are usually considered to be the potential for the application of basic sciences to economic activities, with the proximity of firms to universities and research centers playing a pivotal disseminating role. This view underestimates, however, the importance of research within firms (especially in countries such as Japan and Germany), as well as the dominant role of firms in the creation of new organizing techniques. In this wider perspective, technological opportunities include also the potential in the application of new methods of organizing, with inter-firm relationships as fundamental.

One of the most important results arising from the extensive work done on inter-firm networks in Sweden has been the long-term stability of relationships between buyers and suppliers.[15] It is, in fact, the importance of the ties within an industrial network which forms the basis of Hirschman's (1958) influential statement on the role of forward and backward linkages in economic development. Nor is the pull of these linkages isolated to developing economies.

---

[15] See, for example, Johanson and Mattsson (1987).

Scherer (1982) finds distinctive pattern of externalities among industries in technology flows for the American economy, a pattern that has been observed for other countries.[16] It is this inter-industry diffusion which leads to what Dahmen (1950), in his classic analysis of the growth of the Swedish economy, called 'industry blocks', that is, the spilling-over of technologies into related industries. An important factor in this spilling-over, was the previous experience of entrepreneurs, often from other industries.

As entrepreneurs tend to seek re-employment locally on the basis of accumulated relationships, the slower diffusion of technologies across borders can be partly explained by the immobility of skilled individuals.[17] But the current predominance of organized science over the single inventor suggests a related explanation of the importance of the geographical specificity of relationships among firms and organizations. The studies by von Hippel (1988) have shown that innovations are often responses to requests by suppliers and that users frequently make important contributions to the development of the technology.[18] Moreover, within a country, firms have established over time relationships with research centers, trade associations, and educational institutions (Herrigel, 1992). The higher density of national industrial networks biases these interactions and spillovers towards domestic diffusion.

The study of interdependencies between industries quickly leads to an observation of the interdependencies among countries. Leaving aside the very significant influence of multinational corporations, national boundaries of networks have been eroded by the density of cooperative agreements among firms of different nationalities. Still, most of the evidence on cooperative relationships points to strong national boundaries in industrial networks.[19]

Selection pressures

As discussed earlier, selection pressures and competition are important drivers for innovative efforts, as well as for imitating competitors. To recognize new techniques often requires the pressure of competition in attracting scarce managerial attention (Winter, 1981). As national markets have, historically, been domestic in orientation, the competitive pressures behind selection forces have been largely other national firms. When it spills over borders, competition is, of course, a driver of learning in foreign markets as well.[20] The diffusion of Taylorism and Fordism into France made its first inroads in the auto industries, partly as a response to the competition from Ford's plants located in the United Kingdom (Moutet, 1975). It was only, however, the competition in the extreme guise of World War I which forced France, Germany, and the U.K. to adopt some Taylorist principles.[21]

It is easy to overestimate the importance of network interdependencies relative to the force of competition. In this regard, Dunning's (1986) research on the effects of Japanese investment in the U.K. is especially interesting, for he found that competition was a more important stimulus to the adoption of Japanese technologies in the U.K. than supplier relationships. This effect may possibly reflect the hesitation or inability of Japanese firms to share technology with local suppliers, but it also points to an important role in stimulating new learning.

Identifiability

In most models of learning it is assumed that the individual or organization seeks to adapt its behavior along a gradient in order to approximate some goal. But frequently, the task to be learned is poorly understood and identified. March and Olsen (1975) have especially drawn attention to the phenomenon of 'superstitious' learning, whereby an organization credits an outcome to the wrong cause. Again, it is easy to see why this problem of identification is greater for learning organizational innovations as opposed to a product. The functionality of the product is usually clear; the causal relationship among organizational traits, if it should exist, often involves a high level of complexity.

---

[16] See also the French work on filières, as reviewed by Jacquemin and Rainelli (1984).

[17] Of course, there are many well-known exceptions, such as the forced exodus of the Huguenots from France, as well as the diffusion of British technologies in the nineteenth century to the rest of Europe and to the U.S.

[18] See also Rosenberg (1969) on the diffusion of capital goods in the American economy, and Hounshell (1984) on the movement of employees and development of mass production.

[19] On-going research with Dong-Jac Kim, Weijian Shan and Gordon Walker for the biotechnology and semiconductor industries shows significant numbers of international agreements, but substantially fewer than those within country.

[20] See Benvignati (1982) for a qualified assessment of these effects.

[21] For France, see Moutet (1975) and Boyer (1984); for the U.K., Lewchuk (1987) and Littler (1984); and for Germany, Homburg (1984). An overall view of the diffusion in Europe is described in Devinat (1927).

There is an important implication in an enduring history of competition among national firms. Through continual investments in monitoring and imitating each other, firms develop a capability to understand competitors and to adopt their frontier technologies and best practices. Part of a firm's knowledge is its cumulative understanding of competitors.

The problem of identification is, consequently, more troubling across borders than within a country, for two reasons. The first is best expressed by the concept of signal to noise.[22] Because of their similarity in organizing principles, firms are more sensitive to incremental differences among their domestic competitors. Since there are fewer differences it is easier to identify the factors responsible for differential performance.

Competition by a foreign firm eventually stimulates the search for the underlying advantages. In the absence of previous experience, comparison of the domestic organizational traits against one, or a few, foreign firms raises a surfeit of differences that may account for better performance. To apply a genetic analogy, the number of traits carried out by an organization is in excess of those with performance implications. Only through sufficient observations can the causal mechanisms be inferred, but unlike for domestic competitors, the accumulation of a long history of such observations is often missing.

A second and related reason for the difficulty of identifying the causal linkages responsible for success is the complexity of systemic dependencies. Transferring a just-in-time inventory system requires more than a change in factory procedure; it also involves a high degree of coordination with suppliers. Understanding the elements which need to be changed in tandem is a complex challenge. Deciphering the critical social underpinnings from a complex web of related parties is likely to proceed by trial and error. Because the system (e.g. supplier relationships, institutional connections, labor relations, etc.) which embeds these elements is more different between countries than between firms within the same country, the task of identifying causality is all the more difficult.

Institutional lock-in

The issue of systemic dependence raises the question of the social feasibility of change. These considerations bring into play the contextual reasons for inertia that were omitted from our earlier treatment. Part of the persistence in inferior modes is that, even when the necessary changes are identified, they are not legally or politically feasible to enforce. This perspective was advanced strongly by Veblen (1915) in his analysis of the benefits of the late modernizer, but it has been further developed by Ames and Rosenberg (1963), Kindleberger (1969), and Salter (1966). In this view there is a cost to be an early starter, because technologies change over time, but property rights are slow to adjust.[23]

There is a message in this reasoning. To operate factories or offices at high levels of productivity may take levels of commitment and performance that are not acceptable in all cultures and countries. To adjust to new economic disturbances may also require unacceptable levels of labor mobility, both geographical and occupational. As long as people pursue non-monetary objectives, persistence in economically inferior social arrangements may be a country's choice.

COMPETITIVE CYCLES AND THE CURRENT DILEMMA

No country can attain leadership in absolute advantage without innovation in organizing principles of work. Of course, some countries may converge to (i.e. learn) best practice faster than others. It is this cycle of innovation, within and across country diffusion, and then the introduction of a new innovation (possibly by another country) that generates long-term patterns in competitiveness.

Knowledge is institutionally embedded, but so are the capabilities to learn. The speed by which country capabilities leak across borders is also influenced by the orientation towards learning and the capabilities to learn of individual countries. It has often been noted that the transfer of technology is dependent strongly on the learning capability of the recipient (Hall and Johnson, 1970; Contractor, 1981). The ability to adopt new technologies requires previous investments and education in related areas. Nations do not individually need to be innovative, but their advancement depends upon their institutional capabilities to enjoy the spill-overs from the technological and economic advancement of other countries (Baumol, 1986; Volti, 1980).

---

[22] See Bohn (1988) for an extensive development.

[23] See also the political blockage explanations of Crozier (1964), Olson (1982), Ouchi (1984), and Thurow (1982).

Japanese firms have been frequently cited for their rapid assimilation of foreign technologies, and as exemplifying investments in learning to learn.[24] A wider historical perspective suggests, however, that the roles of imitator and innovator have often been interchanged among countries. It was the imitation of technology by the U.S. and France which was the concern of British manufacturers in the earlier part of the century. In the debates on the imposition of restrictions on the export of capital equipment and on the emigration of workers, Andrew Ure (1835) presented the argument that diffusion was of no significance as long as the Birmingham manufacturers remained on the frontier of technology. In fact, backward engineering, patent infringements, and employment of former workers were all practices in which American firms used to imitate British technology in the nineteenth century (Jeremy, 1981).

There is unlikely to be any simple explanation for this historical pattern of the transition of countries from imitators to innovators as they move into positions of leadership. It might well be that success through the ownership of superior organizing principles forces firms in a country to carry the burden of new technological innovations and shift resources from learning to the creation of new knowledge. However, this change is not likely to be rapid. Because different organizing principles are required for the innovation and imitation of knowledge, organizations cannot easily shift from the former to the latter strategy. The organizational modes of the creation and assimilation of knowledge may not be compatible, not only because the allocation of investment and human resources differs, but also because of differences in the wider societal institutions, such as university-based research (Cole 1985; Westney and Sakakibara, 1985). Imitative capabilities are reflected in the organizational design and institutional relationships within and among firms.

Due to the difference in organizing principles, the transition from emphasizing product innovation to competing on an imitative strategy is equally problematic. Such a dilemma appears to have characterized the United Kingdom and to ail the United States. Finding the U.K. research and development efforts to be extraordinarily high, Peck (1968) suggested that the British government and firms should give greater recognition to an international division of labor in the allocation of research by spending less. A similar analysis and set recommendations are current in the debate on the loss of American competitiveness relative to Japan (Rosenberg and Steinmueller, 1988).

If the characterization of the persistence of the relatively declining country to overemphasize innovation is correct, these trends are not sustainable. Innovating firms in the U.S. will be forced to recognize that the returns to innovative activities are too low. Such a recognition places Japanese firms in an interesting dilemma, for their success poses the possibility of reducing the product innovation efforts of competitors, upon which they have often been dependent. It is not surprising, therefore, that not only have the R&D efforts of Japanese firms increased, but Japanese foreign direct investment has grown by the establishment of the R&D facilities and manufacturing plants in the U.S. and Europe (Kogut and Chang, 1991; Yamawaki, 1990).

This process aids the diffusion of Japanese organizing principles, while potentially placing a larger burden of innovation upon Japan. The positive outlow of U.S. foreign direct investment rode on the superiority of mass production technologies to foreign markets, as suggested in historical accounts (Chandler, 1990; Wilkins, 1974). The outward flows of Japanese investments should be seen as the extension of organizing principles of work to foreign markets. The slowing of Japanese penetration overseas will arise out of the diffusion of its organizing principles, as well as the saturation of those markets served by products best manufactured by these principles.

---

[24] See Clark, Chew and Fujimoto (1987) regarding development times in the auto industry, Mansfield (1988) on the absorption of external technologies, and Westney and Sakakibara (1985) and Imai, Nonaka, and Takeuchi (1985) on the organizational and institutional principles supporting rapid imitation.

There is growing evidence that Japanese organizing principles are diffusing to competitors, especially in the auto industry, with corresponding increases in productivity (MacDuffie and Krafcik, 1990). The response to Japanese competitive pressures in the auto industry has moved from wage-cutting to fundamental changes in the organization of factory work and labor rules, suggesting that a better identification of the sources of Japanese advantage has been achieved (Katz, 1985). Unquestionably, there remains room for further doses of superstitious learning before the organizing principles responsible for the success of Japanese firms are fully understood. But it is also clear that, by the time these differences are diffused, Japan's position in wealth, including its ownership of foreign assets, will be substantially greater than today.

However, the extent of the increase in Japan's wealth will be attenuated by the rate of learning among its foreign competitors. This rate will be influenced not only by diffusion of practices by Japanese subsidiaries, but also by the pulling of investment to Japan and the sourcing of the resident capabilities. There is, then, an important race in progress, namely, between the rate of learning and the evolutionary rate of selection by which non-Japanese firms are displaced from markets. Three separate outcomes are thus possible:

1. Other nations will not adopt new practices, but will compete on older technologies and organizing principles at relatively lower wage and rental (i.e. income) levels.
2. Japanese practices will diffuse by the expansion of the Japanese share of overseas markets, either by exports or direct investment.
3. Firms in other nations will learn the new frontier of best practice.[25]

CONCLUSIONS

The principal contention of the above argument is that competition among developed countries is less a question of *what* product market strategies firms follow, but rather of *how* their productive activities are organized. Of course, there are other important elements. Clearly, issues that affect domestic competitive strategies can be extended to international competition. There are also strategic factors that pertain uniquely to international markets, and that can be analyzed largely in terms of variations in industry structure and consumer tastes among countries.

But variation in industry structure and consumer tastes provides only partial insight into competition within and across countries. Productivity and profitability variations among firms within industries have provoked the recognition that firm-level differences must be treated explicitly in the context of the evolution of industries. The perspective developed above accepts this recognition, with the twist that some of this variation is attributable to the effects of a firm's country of origin. Firms exploit internationally what they have developed domestically, but their capabilities partly reflect their enjoyment of the diffused skills and institutional strengths of their countries of origin. The study of international competition is, in large part, the study of comparative management and societal institutions among countries.

ACKNOWLEDGEMENTS

An earlier version of this paper was presented and published on the occasion of the Crafoord Lectures, held at the University of Lund, 15 November 1989. I would like to acknowledge the comments of Steve Kobrin, Weijian Shan, Gordon Walker, the anonymous referees, and the financial assistance of the German Marshall Fund in supporting the initial research.

REFERENCES

Ames, E. and N. Rosenberg. 'Changing technological leadership and industrial growth', *Economic Journal*, **72**, 1963, pp. 13–31.

---

[25] This result need not lead to convergence, but to alterations in current national practices, as the history of Taylorism and mass production in Europe illustrates.

Armour, H. and D. Teece. ‘Organizational structure and economic performance’. *Bell Journal of Economics*, 9, 1978, pp. 106–122.

Arthur, B. W. ‘Competing technologies, increasing returns, and lock-in by historical events’. *Economic Journal*, 99, 1989, pp. 116–131.

Atkinson, A. B. and J. E. Stiglitz. ‘A new view of technological change’. *Economic Journal*, 79, 1969, pp. 573–578.

Baumol, W. ‘Productivity growth, convergence and welfare: What the long-run data show’. *American Economic Review*, 76, December 1986, pp. 1072–1086.

Bendix, R. *Work and Authority in Industry*. University of California, Berkeley, CA, 1956.

Benvignati, A. ‘The relationship between the origin and diffusion of industrial innovation’. *Economica*, 49, 1982, pp. 313–323.

Bohn, R. ‘Learning in noisy manufacturing environments: An alternative to the learning curve’, working paper no. 88–070, Harvard Business School, 1988.

Boyer, R. ‘Presentation’. In M. de Montmollin and O. Pastre (eds), *Le Taylorisme*, Editions La Decouverte, Paris, 1984, pp. 35–50.

Burenstam Linder, S. *An Essay on Trade and Transformation*, Wiley, New York, 1961.

Cantwell, J. *Technological Innovations and Multinational Corporations*. Basil Blackwell, Cambridge, MA, 1989.

Caves, R. *Economic Analysis and Multinational Enterprises*, Cambridge University Press, Cambridge, 1982.

Chandler, A. *Scale and Scope. The Dynamics of Industrial Capitalism*. Harvard University Press, Cambridge, MA, 1990.

Clark, K., W. Bruce Chew and T. Fujimoto. ‘Product development in the world auto industry’, *Brookings Papers on Economic Activity*, 3, 1987, pp. 729–782.

Cole, R. ‘The macropolitics of organizational change: A comparative analysis of the spread of small-group activities’, *Administrative Science Quarterly*, 30, 1985, pp. 560–585.

Contractor, F. *International Technology Licencing: Compensation, Costs and Negotiation*, Lexington Press, Lexington, MA, 1981.

Coriat, Benjamin. *L'atelier et le Chronometre*, Christian Bourgois Editeur, Paris, 1979.

Crozier, M. *The Bureaucratic Phenomenon*, University of Chicago Press, Chicago IL, 1964.

Cusumano, M. *The Japanese Automobile Industry*, Council on East Asiān Studies, Harvard University Press, Cambridge, MA, 1985.

Cyert, R. and J. March. *A Behavioral Theory of the Firm*, Prentice Hall, Englewood Cliffs, NJ, 1963.

Daft, R. L. ‘A dual-core model of organizational innovation’, *Academy of Management Journal*, 21, 1978, pp. 193–210.

Dahmen, E. *Svensk Industriell Foretagarverksamhet*, Industrins Utredningsinstitut, Stockholm, 1950.

Damanpour, F. and W. M. Evan. ‘Organizational innovation and performance: The problem of organizational lag’, *Administrative Science Quarterly*, 29, 1984, pp. 392–409.

David, P. *Technical Choice, Innovation and Economic Growth: Essays on American and British Experience in the Nineteenth Century*, Cambridge University Press, New York, 1975.

Davies, M. *Woman's Place is at the Typewriter: Office Work and Office Workers, 1870–1930*, Temple University Press, Philadelphia, PA, 1982.

Deane, P. *The First Industrial Revolution*, Cambridge University Press, Cambridge, 1975.

Deardorff, A. ‘Testing trade theories and predicting trade flows’. In R. W. Jones and P. B. Kenen (eds), *Handbook of International Economics* Elsevier Science Publishers B.V., Amsterdam, 1984, pp. 467–517.

Devinat, P. *Scientific Management in Europe*, ILO Studies and Reports, Series B, no. 17, Geneva, 1927.

Dohse, K., U. Juergens and T. Malsch. ‘From “Fordism” to “Toyotism”? The social organization of the labor process in the Japanese automobile industry’, *Politics and Society*, 14, 1985, pp. 115–146.

Dore, R. *British Factory, Japanese Factory: The Origins of Diversity in Industrial Relations*, University of California Press, Berkeley, CA, 1973.

Dosi, G. ‘Technological paradigms and technological trajectories: A suggested interpretation of the determinants and directions of technical change’, *Research Policy*, 11, 1982, pp. 147–163.

Dunning, J. *American Investment in British Manufacturing Industry*, George Allen & Unwin, London, 1958.

Dunning, J. *Japanese Participation in British Industry: Trojan Horse or Catalyst for Growth*, Croom-Helm, Dover, NH, 1986.

Dunning, J. ‘The globalization of firms and the competitiveness of countries’. In *Globalization of Firms and the Competitiveness of Nations*, Crafoord Lecture, University of Lund, Lund, 1990, pp. 9–58.

Freeman, C. *The Economics of Industrial Innovation*, MIT Press, Cambridge, MA, 1982.

Freeman, C. *Technology Policy and Economic Performance*, Pinter, London, 1987.

Fridenson, P. ‘Un tournant taylorien de la societé Francaise (1904–1908)’, *Annales ESC*, 5, 1987, pp. 1031–1060.

Fruen, M. ‘Cooperative structure competitive strategies: The Japanese enterprise system’, Manuscript, INSEAD, 1989.

Griliches, Z. ‘Productivity R&D and basic research at the firm level in the 1970s, *American Economic Review*, 76, 1986, pp. 141–144.

Hall, G. and R. Johnson. ‘Transfers of United States aerospace technology to Japan’. In R. Vernon (ed.), *The Technology Factor in International Trade*, Columbia University Press, New York, 1970, pp. 305–358.

Herrigel, G. ‘The polities of large firm relations with industrial districts: A collision of organizational fields in Baden Wuerternberg’. In B. Kogut (ed.), *Country Competitiveness and the Organizing of Work and Technology*, Oxford University Press, Oxford (forthcoming, 1992).

Hirschman, A. *Strategy of Economic Development*.

Yale University Press. New Haven. CT. 1958.

Homburg. H. 'Le Taylorisme et la Rationalisation de l'organization du travail en Allemagne (1918–1939). In M. de Montmollin and O. Pastre (eds). *Le Taylorisme*. Editions la Decouverte. Paris. 1984. pp. 99–114.

Hounshell. D. A. *From the American System to Mass Production 1800–1932*. Johns Hopkins University Press. Baltimore. MD. 1984.

Imai. K.. I. Nonaka and H. Takeuchi. 'Managing the new product development process: How Japanese companies learn and unlearn'. In K. Clark. R. Hayes and C. Lorenz (eds). *The Uneasy Alliiance Managing the Productivity–Technology Dilemma*. Harvard Business School Press. Boston. MA. 1985. pp. 336–376.

Jacquemin. A. and M. Rainelli. 'Filiere de la Nation et Filiere de l'Entreprise.' *Revue Economique*. **26**. 1984. pp. 379–392.

Jeremy. D. J. *Transatlantic Industrial Revolution: The Diffusion of Textile Technologies Between Britain and America 1790–1830s*. MIT Press. Cambridge. MA. 1981.

Johanson. J. and L. G. Mattsson. 'Internalisation in industrial networks—a network approach'. In N. Hood and J. E. Vahlne (eds). *Strategies in Global Competition*. Croom-Helm. London. 1987. pp. 287–314.

Katz. H. *Shifting Gears: Changing Labor Relations in the U.S. Automobile Industry*. MIT Press. Cambridge. MA. 1985.

Kindleberger. C. P. *American Business Abroad*. MIT Press. Cambridge. MA. 1969.

Kogut. B. 'Country patterns in international competition: Appropriability and oligopolistic agreement'. In N. Hood and J. E. Vahlne (eds). *Strategies in Global Competition*. Croom-Helm. London. pp. 315–340.

Kogut. B. 'The permeability of borders and the speed of learning among countries'. In *Globalization of Firms and the Competitiveness of Nations*. Crafoord Lecture. University of Lund. Lund. pp. 59–96.

Kogut. B. and S. J. Chang. 'Technological capabilities and Japanese foreign direct investment in the United States'. *Review of Economics and Statistics* (forthcoming. 1991).

Kogut. B. and D. Parkinson. 'The diffusion of American organizing principles to Europe'. In B. Kogut. (ed.). *Country Competitiveness and the Organizing of Work and Technology*. Oxford University Press. Oxford (forthcoming. 1992).

Kogut. B. and U. Zander. 'Knowledge of the firm and the international replication of technology'. mimeo. Wharton School of Business and Institute of International Business. Stockholm School of Economics. 1990.

Krafcik. J. 'Triumph of the lean production system'. *Sloan Management Review*. **29**. 1988. pp. 41–52.

Laux. J. *In First Gear. The French Automobile Industry to 1914*. McGill-Queen's University Press. Montreal. 1976.

Levin. R.. A. Klevorick. R. Nelson and S. Winter. 'Appropriating the returns from industrial research and development'. *Brookings Papers on Economic Activity*. 3. 1987. pp. 783–820.

Lewchuk. W. *American Technology and the British Vehicle Industry*. Cambridge University Press. Cambridge. 1987.

Lincoln. J.. M. Hanada and K. McBride. 'Organizational structures in Japanese and U.S. manufacturing'. *Administrative Science Quarterly*. 31. 1986. pp. 338–364.

Lippman. S. A. and R. P. Rumelt. 'Uncertain imitability: An analysis of interfirm differences in efficiency under competition'. *Bell Journal of Economics*. 13, 1982. pp. 418–438.

Littler. C. 'L'essor du Taylorisme et de la rationalisation du travail dans l'industrie anglaise (1880–1939)'. In M. de Montmollin and O. Pastre (eds). *Le Taylorisme*. Editions La Decouverte. Paris. 1984. pp. 83–97.

MacDuffe. J. P. and J. Krafcik. 'Integrating technology and human resources for high performance manufacturing: Evidence from the international auto industry'. mimeo. International Motor Vehicle Program. MIT. 1990.

Maddison. A. *Phases of Capitalist Development*. Oxford University Press. Oxford. 1982.

Maier. C.S. 'Between Taylorism and technocracy: European ideology and the vision of industrial productivity in the 1920s'. *Journal of Contemporary History*. 5. 1970. pp. 27–61.

Malerba. F. 'The Italian system of innovations'. Presented at the workshop on 'The Organization of Work and Technology: Implications for International Competitiveness'. Brussels. June 1990.

Mansfield. E.M. 'How rapidly does new industrial technology leak out?' *Journal of Industrial Economics*. 35. 1985. pp. 217–223.

Mansfield. E.M. 'Speed and cost of industrial innovation in Japan and the United States'. *Management Science*. 34. 1988. pp. 1157–1168.

March. J. and J. Olsen. 'The uncertainty of the past: Organizational learning under ambiguity'. *European Journal of Political Research*. 3. 1975. pp. 147–171.

Mason. M. 'Access denied. American direct investment in Japan'. mimeo. Harvard University. 1990.

Maurice. M.. F. Sellier and J. J. Silvestre. 'The search for a societal effect in the production of company hierarchy: A comparison of France and Germany'. In P. Osterman (ed.). *Internal Labor Markets*. MIT Press. Cambridge. MA. 1984. pp. 231–270.

McInnes. J.M. 'Financial control systems for multinational operations'. *Journal of International Business Studies*. 1. 1971. pp. 11–28.

Moutet. A. 'Les origines du systeme de Taylor en France. Le point de vue patronal (1907–1914)'. *Le Mouvement Social*. 93. 1975. pp. 15–49.

Nelson. D. *Frederick W. Taylor and the Rise of Scientific Management*. University of Wisconsin. Madison. WI. 1980.

Nelson. R. R. 'Research on productivity growth and differences'. *Journal of Economic Literature*. 19. 1981. pp. 1029–1064.

Nelson, R. R. and S. G. Winter. 'In search of useful theory of innovation', *Research Policy*, 6, 1977, pp. 36–76.

Nelson, R. R. and S. G. Winter. *An Evolutionary Theory of Economic Change*. Harvard University Press, Cambridge, MA, 1982.

Olson, M. *The Rise and Decline of Nations: Economic Growth, Stagflation, and Social Ridigities*. Yale University Press, New Haven, CT, 1982.

Ouchi, W. M. *Theory Z*. Addison-Wesley, Reading, MA, 1981.

Ouchi, W. M. *The M-Form Society: How American Teamwork can Recapture the Competitive Edge*. Addison-Wesley, Reading, MA, 1984.

Pavitt, K. 'International patterns of technological accumulation'. In N. Hood and J. E. Vahne (eds), *Strategies in Global Competition*. Croom-Helm, London, 1989, pp. 126–157.

Pavitt, K. and L. L. G. Soete. 'International differences in economic growth and the international location of innovation. In H. Giersch (ed.), *Emerging Technologies: Consequences for Economic Growth, Structural Change and Employment*, J. C. B. Mohn, Tubingen, 1982, pp. 105–133.

Peck, M. J. 'Science and Technology'. In R. Caves *et al.* (eds), *Britain's Economic Prospects*, Brookings Institute, Washington, DC, 1968, pp. 448–486.

Piore, M. and C. Sabel. *The Second Industrial Divide*, Free Press, New York, 1984.

Porter, M. *The Competitiveness of Nations*. Free Press, New York, 1990.

Rogers, E. *Diffusion of Innovations*, Free Press, New York, 1971.

Rosenberg, N. 'The direction of technological change: Inducement mechanisms and focusing devices', *Economic Development and Cultural Change*, **18**, 1969, pp. 1–24.

Rosenberg, N. and W. E. Steinmueller. 'Why Americans are such poor imitators'. *American Economic Review. Papers and Proceedings*, **78**, pp. 229–234.

Rostas, L. *Comparative Productivity in British and American Industry*, Cambridge University Press, Cambridge, 1948.

Rumelt, R. P. 'Towards a strategic theory of the firm'. In R. Lamb (ed.), *Competitive Strategic Management*. Prentice-Hall, Englewood Cliffs, NJ, 1984, pp. 566–570.

Salter, W. *Productivity and Technical Change*, Cambridge University Press, Cambridge, 1966.

Scherer, F. M. 'Firm size, market structure, opportunity and the output of patented inventions', *American Economic Review*, **55**, 1965, pp. 1097–1123.

Scherer, F. M. 'Interindustry technology flows in the United States'. *Research Policy*, **11**, 1982, pp. 227–245.

Servan-Schreiber, J.-J. *The American Challenge*, Athenaeum, New York, 1969.

Smith, M. R. *Harpers Ferry Armory and the New Technology. The Challenge of Change*, Cornell University Press, Ithaca, NY, 1977.

Stiglitz, J. E. 'Learning to learn, localized learning and technological progress'. In P. Dasgupta and P. Stonema (eds), *Economic Policy and Technological Performance*. Cambridge University Press, New York, 1987, pp. 125–133.

Stinchcombe, A. 'Social structure and organizations'. In J. G. March (ed.), *Handbook of Organizations*, Rand-McNally, Chicago, IL, 1965, pp. 142–193.

Swendenborg, B. *The Multinational Operations of Swedish Firms*, Industrial Institute for Economic and Social Research, Stockholm, 1979.

Teece, D. 'Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy'. In D. Teece (ed.), *The Competitive Challenge–Strategies for Industrial Innovation and Renewal*, Ballinger, Cambridge, MA, 1987, pp. 185–220.

Thurow, L. C. *The Zero-Sum Society. Distribution and the Possibilities for Economic Change*, Penguin Books, New York, 1982.

Tilton, J. *The International Diffusion of Technology. The Case of Semiconductors*, Brookings Institution, Washington, DC, 1971.

Ure, A. *The Philosophy of Manufacturers*, London, 1835.

Velben, T. *Imperial Germany and the Industrial Revolution*, Macmillan, New York, 1915.

Volti, R. 'The absorption and assimilation of acquired technology'. In R. Baum (ed.), *China's Four Modernizations*, Westview Press, Boulder, CO, 1980, pp. 179–201.

von Hippel, E. *The Sources of Innovation*, Oxford University Press, Oxford, 1988.

Westney, E. *Imitation and Innovation. The Transfer of Western Organizational Patterns to Meiji Japan*, Harvard University Press, Cambridge, MA, 1989.

Westney, E. and D. Sakakibara. 'Competitive study of the training careers and organization of engineers on the computer industry in Japan and the United States', mimeo, MIT-Japan Science and Technology Program, 1985.

Wilkins, M. *The Maturing of Multinational Enterprise: American Business Abroad from 1914 to 1970*, Harvard University Press, Cambridge, MA, 1974.

Winter, S. 'Attention allocation and input proportions', *Journal of Economic Behavior and Organization*, 2, 1981, pp. 31–46.

Winter, S. 'Knowledge and competence as strategic assets'. In D. Teece (ed.), *The Competitive Challenge—Strategies for Industrial Innovation and Renewal*, Ballinger, Cambridge, MA, 1987, pp. 159–184.

Yamawaki, H. 'Location decisions of Japanese multinational firms in European manufacturing industries', working paper, Wissenschaftszentrum, Berlin, 1990.

74

WHARTON REPROGRAPHICS

LIPF
DOC

The geographical sources of the competitiveness of firms: some results of a new survey

John H. Dunning[*]

This article presents some results of a new field survey, designed to assess the extent to which the executives of some of the largest industrial corporations in the world perceived how they had augmented their global competitive advantages as a result of their foreign direct investments. Among other interesting findings, the survey revealed that the more transnational a firm was, the more it derived its competitive advantages from its foreign affiliates; while, over the last decade or so, an increasing proportion of these competitive advantages were obtained from this source.

1. Introduction

This article presents some of the results of a new field survey of the geographical sources of firm-specific competitiveness as perceived by executives of 144 of the world's 500 largest industrial firms,[1] all of which are transnational corporations (TNCs). In 1993, the sample firms were responsible for 39 per cent of the global sales and nearly two fifths of the foreign direct investment (FDI) of the largest 500 industrial corporations. For some industries (e.g. motor vehicles, petroleum refining, tobacco, electronics and computers) and for some countries (e.g. Japan, Sweden and the United Kingdom), the coverage of the survey is particularly good; for others, it is less so; but given the country and industry groups used here, it is entirely acceptable. Full details of the sample and the way in which the survey was organized and carried out are presented in the appendix.

---

[*] State of New Jersey Professor of International Business, Rutgers University, Newark, New Jersey, United States. The author is much indebted to Sarianna Lundan of Rutgers University for research assistance and to the anonymous referees for helpful suggestions.

[1] As identified by *Fortune* magazine, 25 July 1994.

The purpose of the field survey, conducted in the second half of 1994, was to obtain the opinions of senior business executives[2] on the geographical origin of the kind of firm-specific competitive advantages or core competencies—sometimes called ownership-specific advantages of firms—identified by the literature.[3] There has been much debate in recent years over whether or not the competitive advantages of firms—some of which may be internalized by the firms themselves—stem from the location-bound characteristics of their home countries, or whether or not, as FDI and strategic alliances have become aimed increasingly at acquiring created assets[4] and as firms have become more transnational,[5] firms' competitive advantages—particularly knowledge of one kind or another—are being sourced increasingly from outside their home countries. In other words, to put it in Michael Porter's terminology, whether or not firms source from "multiple home bases" (Porter, 1997).

The concept of the "double" diamond of competitive advantage, as developed by Alan Rugman and his colleagues over the past few years,[6] is used here as the analytical basis. Essentially, the Rugman approach extends the concept of Porter's "single" diamond of competitive advantage (first set out in Porter, 1990) that argues that a firm's competitive advantages are essentially a function of the domestic economic environment in which it operates. Rugman and others, including the present author Dunning (1993b), have argued that the deepening structural integration of the world economy and the burgeoning of alliance capitalism (Dunning, 1995) are widening the geographical scope for creating or augmenting firm-specific competencies and learning experiences. Any attempt to identify the geographical sources of such advantages must embrace the diamonds of other countries, particularly those with which the home-country firms have the most dealings by way of trade, FDI and non-equity cooperative ventures.

---

[2] Mainly Vice-Presidents or Directors of Foreign Operations.

[3] For a summary of the competitive advantages identified by economists, see, for example, Dunning (1993a, chap. 4), and for those identified by business strategists, see, for example, Porter (1990) and Peteraf (1993). For recent analyses of knowledge sourcing by TNCs outside their national boundaries, see Almeida (1996), Kuemmerle (1995 and 1996) and Pearce and Singh (1991 and 1992).

[4] The aim of strategic asset-acquiring FDI is to gain access to assets that protect or augment the acquiring firm's competitive advantages and/or lessen those of its rivals. In the language of the eclectic paradigm, firms engage in FDI not to exploit existing ownership (O) advantages, but to gain new advantages, which, when deployed with their existing O advantages, help sustain or further their global competitive competencies.

[5] As documented, for example, in various *World Investment Reports* (UNCTAD-DTCI, 1995 and 1996). The most recent United States data (Mataloni, 1995) show that the foreign affiliate sales of non-bank United States parent firms were 45 per cent of their parent sales in 1993. This figure compares with a corresponding 1983 figure of 37 per cent.

[6] See the special issue of *Management International Review* edited by Alan Rugman (1993), and Rugman, Van den Broeck and Verbeke (1995).

This field study aims at offering some new evidence to that recorded by a number of country case-studies[7] which, in the main, have confirmed the Rugman/Dunning hypothesis. The uniqueness of the present contribution is sixfold:

- The information is obtained directly from firms.
- The competitive advantages identified in the literature are classified into groups using as the framework the four components of Porter's diamond.
- Its findings are related to a number of critical firm-specific variables, viz. size of firm, degree of transnationality, country (or region) of origin and industry.
- Those foreign countries that provide the principal access to non-domestic competitive advantages that complement those offered by the home country are identified.
- The sample firms are ranked by the significance of the three main modes of foreign involvement used by these firms in acquiring competitive advantages from a foreign location.
- The perceptions of the sample firms about the positive or negative influences of different components of home-government policy on their global competitiveness are set out.

To keep this article within manageable dimensions, a threefold classification of industries is adopted, based upon their degree of research and development (R & D) intensity in the United States in 1992. Thus, high-technology (HT) industries are defined as those in which the average R & D expenditure as a percentage of sales was at least 4 per cent, and in which scientists and engineers employed in R & D as a percentage of total employment were 2 per cent or more.[8] Medium-technology (MT) industries are defined as those in which the corresponding ratios vary between 2 per cent and 3.9 per cent, and 1 per cent and 1.9 per cent, respectively; and low-technology (LT) industries are defined as those that embrace those ratios under 2 per cent and under 1 per cent, respectively. As to home country or region of origin, five main groups, grouped according to their size and economic profiles, are considered here. These are, respectively, "larger" European countries,[9] "smaller" European countries,[10] the United States, Japan, developing economies,[11] and other countries.[12]

---

[7] Including Australia, Canada, Denmark, Republic of Korea, Mexico and New Zealand. See Rugman (1993).

[8] More specifically, HT industries include pharmaceuticals; MT sectors include industrial equipment, motor vehicles, petroleum refining, soap and cosmetics; and LT sectors include beverages, building materials, food, metal products, paper, publishing and printing, rubber and plastics, textiles and tobacco.

The firms in the sample are also grouped by their degrees of transnationality. This was obtained by averaging out (for each firm) the percentages of their global assets and global employment[13] accounted for by their foreign affiliates,[14] and then classifying firms into four groups on the basis of the degree of transnationality, viz. under 15 per cent (low), 15-29 per cent (medium-low), 30-59 per cent (medium-high) and 60 per cent and over (high).

Finally, although all of the sample firms were large companies, their global sales in 1993 varied from US$ 0.22 billion to US$ 138.2 billion. For the purposes of the analysis, firms were categorized into four groups according to size; medium-size (M) firms, defined as those with sales of under US$ 5 billion in 1993; large (L) firms, with sales between US$ 5 billion and US$ 24.9 billion in that year; very large (VL) firms, with sales between US$ 25 billion and US$ 49 billion; and mega (Mg) firms, with US$ 50 billion and over in sales.

In future research, an econometric analysis of the relationship between these four contextual variables and the competitive advantages of firms will be undertaken, since, with few exceptions, data exist for each of the 144 firms. For now however, the analysis is confined to identifying and exploring a number of possible bi-variant (and occasionally tri-variant) relationships between the sources of competitiveness and firm- and/or country-specific characteristics, which—at the very least—may help point the way to more rigorous statistical evaluation.

---

[9] These include France, Germany, Italy, Spain and the United Kingdom. Two United Kingdom/Netherlands-based TNCs are also included as originating from this group of countries.

[10] These include Belgium, Denmark, Finland, the Netherlands, Sweden and Switzerland.

[11] These include Brazil, Chile, Hong Kong, the Republic of Korea, Mexico and Taiwan Province of China.

[12] These include Canada, Australia and New Zealand.

[13] Or one of these when data on either employment or assets were not available. Additionally, for some firms, sales data from Worldscope/Disclosure were used.

[14] As defined by the TNCs themselves.

2. The sample firms: some details

Tables 1 to 3 set out some details about the characteristics of the sample firms. In general, they mirror the existing knowledge about the industrial and geographical composition of TNCs, as well as the size and degree of transnationality as it has been set out in various publications, for example, *World Investment Reports 1991-1996* (UNCTC, 1991 and 1992; UNCTAD-DTCI, 1993, 1994, 1995, 1996) and *World Investment Directory* (4 vols.,

Table 1. **The industrial distribution and extent of transnationality of the sample firms, 1993**

(*Billions of dollars and percentage*)

Industry	Number of firms	Sales (*Billions of dollars*)	Sales (*Percentage*)	Extent of transnationality[a] (*Percentage*) Assets	Employment	R & D
High technology (HT), of which:	53	892.0	36.0	38.2	39.8	22.7
Aerospace	5	73.1	2.9	6.5	6.2	2.0
Chemicals and pharmaceuticals	19	215.9	8.7	51.3	54.2	26.6
Computers	9	213.9	8.6	28.9	28.7	9.6
Electronics	20	389.1	15.7	38.8	39.2	27.5
Medium technology (MT), of which:	34	1 152.2	46.5	39.6	39.4	15.4
Industrial equipment	8	73.8	3.0	28.0	39.1	12.7
Motor vehicles	12	580.0	23.4	32.2	35.5	7.5
Petroleum refining	14	498.4	20.1	49.7	42.9	24.3
Low technology (LT), of which:	57	434.5	17.5	41.7	45.5	28.5
Food, drink and tobacco	20	244.3	9.9	50.4	58.1	32.5
Paper	10	39.0	1.6	22.4	20.9	8.1
Building materials	5	16.4	0.7	49.6	46.4	39.7
Metals and metal products	14	107.3	4.3	35.0	38.0	24.6
Other industries	8	27.6	1.1	53.5	59.6	45.6
Total	144	2 478.7	100.0	40.0	41.8	23.3

Source: Author's estimates, based on the survey of sample firms.

[a] Measured as the percentage share of total assets, employment and R & D of the respondent firms accounted for by their foreign affiliates.

Table 2. The distribution of the sample firms, by region or country of origin and extent of transnationality, 1993

(Billions of dollars and percentage)

Region/country	Number of firms	Sales (*Billions of dollars*)	Sales (*Percentage*)	Extent of transnationality (*Percentage*) Assets	Employment	R & D
"Large" European countries, of which:	44	922.8	77.2	49.1	49.9	32.0
Germany	13	359.5	14.5	39.7	46.0	23.1
United Kingdom and United Kingdom/Netherlands	23	400.3	16.1	57.3	57.9	38.6
Other countries	8	163.0	6.6	28.8	32.1	27.8
"Small" European countries, of which:	21	164.5	6.6	55.2	56.8	32.7
Sweden	7	32.0	1.3	70.0	70.3	53.0
Switzerland	5	68.9	2.8	58.0	61.8	33.2
Other countries	9	63.6	2.6	46.1	43.6	19.7
United States	31	710.2	28.6	36.1	34.5	14.7
Japan	26	606.3	24.5	19.5	28.7	6.0
Developing countries	14	42.5	1.7	27.2	25.4	15.6
Other countries	8	32.4	1.3	30.5	30.6	18.3
Total	144	2 478.8	100.0	40.0	41.8	23.3

Source: Author's estimates, based on the survey of sample firms.

UNCTAD-DTCI, 1992-1994). For the most part, too, the data set out in the tables are self-explanatory. However, three points should be highlighted:

- Although 82 per cent of the sales of the sample firms were within the high- or medium-technology industries, the degree of technological intensity did not appear to affect significantly the extent of their transnationality. There were, indeed, as wide differences among industries within the three broad technological intensity groupings identified in table 1 (compare, for example, aerospace with chemicals) as between these groupings.

- The extent to which TNCs undertake R & D activities outside their home countries is, on average, about one half of the share of their foreign assets and employment in their respective totals, although this

Table 3. The distribution of the sample firms, by size and extent of transnationality, 1993

(Billions of dollars and percentage)

Size by value of sales	Number of firms	Sales (*Billions of dollars*)	Sales (*Percentage*)	Extent of transnationality (*Percentage*) Assets	Employment	R & D
Medium (< 5 billion dollars)	56	141.6	5.7	40.6	39.7	20.1
Large (5-24.9 billion dollars)	58	706.2	28.5	37.3	41.1	26.5
Very Large (25-49.9 billion dollars)	16	561.5	22.6	46.1	45.0	29.0
Mega (> 50 billion dollars)	14	1 069.4	43.1	42.3	49.5	12.8
Total	144	2 478.8	100.0	40.0	41.8	23.3

Source: Author's estimates, based on the survey of sample firms.

does vary between groupings (table 1). Table 2, however, shows that the R & D transnationality ratio ranges from 6 per cent in the case of Japanese TNCs to 53 per cent in the case of Sweden-based TNCs.

- Table 3 shows that the size of a firm does not appear to be a significant influence on the degree of transnationality.[15] Somewhat surprisingly, however, the mega-firms—although the most transnational in terms of their sales and assets—were the least transnational in terms of their R & D expenditures.

3. The geographical sources of competitiveness

Tables 4 to 7 present the evaluation of the respondents of the sample firms of the extent to which they believe they currently derive a number of identified specific competitive advantages as a direct result of their FDI and/or strategic alliances with foreign firms. The respondents were asked to use a scale of 1 to 7 in their evaluations, with a rank of 1 indicating that their competitive advantages were *entirely* derived from the resources, capabilities, markets and inter-firm rivalries specific to their home countries. (Conversely, a rank of 7 would indicate that such advantages were derived entirely from the location-specific attributes of foreign countries.) A rank of 4 would suggest that, in the respondent's opinion, the origin of the competitive advantages stemmed equally from the home country and foreign locations.[16]

---

[15] However, it should be remembered that all the firms in this sample are large, compared with the great majority of the 38,000 parent firms identified by UNCTAD in the *World Investment Reports 1995* and *1996*.

(a) *Sector-specific differences*

Table 4 considers four groups of competitive advantages that correspond broadly to Porter's fourfold diamond of a country's competitive advantage.[17] Three of these four groups have been subdivided further into a number of more specific advantages, identified by other scholars. In table 4 the sample firms are also classified into high, medium- and low-intensity groupings. Economic theory would suggest that firms should seek to acquire competitive assets from those locations that provide these assets at the lowest cost and with the greatest security.

Essentially, the Porter hypothesis is that, independently of their size, industry, country of origin and degree of transnationality, firms will create, or obtain, their competitive advantages (e.g. innovatory capacity) in, or from, their home countries. The contrasting hypothesis, set out by Rugman, Dunning and other scholars researching into the behaviour of TNCs, is that, as firms become more transnational and globally integrate as regards their value-added activities, they are likely to derive an increasing proportion of their competencies from outside national boundaries. And, indeed, firms may deliberately seek out foreign assets that they perceive to augment, or be complementary to, these competencies.

The data in table 4 reveal a mixed picture, although, taken as a whole, they do suggest that the sample firms derive an important part of their competitive advantages—on average between 40 per cent and 50 per cent[18]—from their presence in foreign countries, either by way of FDI or strategic

---

[16] The comparison in this exercise is between advantages thought to emanate from the portfolio of assets located in the home country and that in all other countries.

[17] Viz. factor conditions, demand conditions, firm strategy, structure and rivalry, and related and supporting industries (Porter, 1990).

[18] Each ranking is being expressed as a percentage of 7 (a ranking that would suggest that all firms' advantages are derived from outside the national boundaries).

Table 4. The sourcing of competitive advantages, classified by the technological intensity of the sample firms, 1993

(Mean value)

	All	High technology	Medium technology	Low technology
Group 1 **Access to resources and assets**	3.54 (1.39)	3.46 (1.20)	3.48 (1.03)	3.65 (1.71)
(i) Natural resources	4.24 (1.70)	3.67 (1.36)	4.44 (1.67)	4.62 (1.88)
(ii) Unskilled labour	3.98 (1.70)	4.10 (1.55)	3.79 (1.65)	3.99 (1.89)
(iii) Skilled and professional labour	2.98 (1.27)	2.98 (1.06)	2.76 (1.17)	3.12 (1.50)
(iv) Innovatory capacity	2.88 (1.40)	2.75 (1.16)	2.71 (1.17)	3.11 (1.70)
(v) Organizational capacity	3.12 (1.32)	3.21 (1.13)	2.88 (1.04)	3.18 (1.61)
(vi) Managerial expertise	3.24 (1.42)	3.19 (1.32)	3.12 (1.37)	3.38 (1.56)
(vii) Relational skills	3.75 (1.62)	4.00 (1.64)	3.41 (1.33)	3.71 (1.73)
Group 2 **Consumer demand**	3.94 (1.59)	4.06 (1.36)	3.37 (1.50)	4.15 (1.81)
(i) Upgrading of product quality	3.31 (1.37)	3.40 (1.36)	2.94 (1.14)	3.44 (1.49)
(ii) Making for more product innovation	3.44 (1.45)	3.40 (1.26)	3.06 (1.50)	3.71 (1.55)
Group 3 **Inter-firm competition/rivalry**	4.60 (1.67)	4.68 (1.61)	4.56 (1.65)	4.55 (1.75)
Group 4 **Linkages with foreign or domestic firms and institutions**	4.10 (1.37)	4.19 (1.32)	3.68 (1.12)	4.29 (1.52)
(i) Related firms (agglomerative economies)				
(ii) Universities and other research institutions	3.29 (1.30)	3.27 (1.25)	3.21 (1.17)	3.38 (1.43)

Source: Author's estimates, based on the survey of sample firms.

NOTE: Unless otherwise indicated, in this and all subsequent tables the figures reported are mean responses with standard deviations in parentheses.

alliances. The access provided by a foreign location (compared with a domestic location) to natural resources (including unskilled labour), linkages with suppliers, industrial competitors and other foreign producers, and the benefits of larger markets and more stringent consumer demands, were all ranked particularly high (3.44 or above). By contrast, and in line with received theory and empirical evidence on innovatory activity, the sample firms perceived that their domestic operations and/or the indigenous resources and capabilities of their home countries continued to provide the main source of competitiveness—especially in terms of technological capacity and skilled professional manpower.

Predictably, there were some differences in the perceptions of firms depending on the technological intensity of their main activities. Most noticeably, LT firms claimed to obtain a higher proportion of their created assets (group 1 [iii-vii]) from foreign sources than did MT or HT firms, particularly so in the case of innovatory capacity. *Inter alia*, this finding tends to support the proposition that at least the part of the competitive advantages of firms that is obtained from foreign sources is likely to be different from (and complementary to) the part that is obtained from domestic sources. Nevertheless, the evidence suggests strongly that, for each of the advantages identified, a multiple (or at least a dual) location of value-added activities was perceived to yield positive gains.[19] Thus, it seems reasonable to conclude that, in the opinion of the leading industrial companies, FDI and/or cross-border strategic alliances *do* provide access to significant competitive advantages.[20]

(b) *Country-specific differences*

Table 5 reclassifies the data set out in table 4 by the source region or country of the sample firms. The data reveal few significant differences in the perceptions of the sample firms. As might be reasonably predicted, firms from high-wage economies, e.g. the United States and smaller European countries,[21] ranked the foreign sourcing of unskilled labour relatively higher than firms from low-wage economies, particularly developing countries. Likewise, and consistent with the principle of comparative advantage, was the above-average reliance of Japanese firms on foreign natural resources, and the below-average reliance of other developed-country firms on such

---

[19] A future article will attempt to relate the performances of the sample firms to the sources of their competencies.

[20] Of course, it is recognized that these data do not, in themselves, indicate whether the overall competitive position of the respondent firms would have been better or worse, if the FDI had *not* taken place, or the strategic alliances had *not* been formed.

[21] Notably Sweden and Switzerland.

Table 5. The sourcing of competitive advantages, classified by region or country of origin of the sample firms, 1993

(Mean value)

	All	"Large" European countries	"Small" European countries	United States	Japan	Other developed countries	Developing countries
Group 1 **Access to resources and assets**	3.54 (1.39)	3.80 (1.25)	3.90 (1.41)	3.61 (1.43)	3.21 (1.35)	2.63 (1.30)	3.21 (1.63)
(i) Natural resources	4.24 (1.70)	4.26 (1.58)	3.65 (1.81)	4.03 (1.61)	4.87 (1.63)	3.29 (2.14)	4.86 (1.66)
(ii) Unskilled labour	3.98 (1.70)	3.81 (1.63)	4.10 (1.77)	4.37 (1.50)	3.75 (1.80)	4.00 (1.91)	3.92 (2.10)
(iii) Skilled and professional labour	2.98 (1.27)	3.18 (1.22)	3.30 (1.45)	3.27 (1.15)	2.46 (1.26)	3.00 (1.26)	2.21 (1.01)
(iv) Innovatory capacity	2.88 (1.40)	2.95 (1.31)	3.38 (1.56)	2.94 (1.39)	2.23 (1.34)	3.29 (1.25)	2.79 (1.37)
(v) Organizational capacity	3.12 (1.32)	3.30 (1.37)	3.43 (1.12)	3.35 (1.05)	2.85 (1.54)	2.43 (1.27)	2.43 (1.28)
(vi) Managerial expertise	3.24 (1.42)	3.30 (1.49)	3.86 (1.39)	3.29 (1.10)	3.12 (1.70)	3.14 (0.90)	2.36 (1.22)
(vii) Relational skills	3.75 (1.62)	3.98 (1.56)	4.43 (1.69)	3.90 (1.35)	3.46 (1.70)	3.29 (1.70)	2.43 (1.28)
Group 2 **Consumer demand**	3.94 (1.59)	4.39 (1.64)	4.22 (1.73)	3.78 (1.53)	2.96 (1.19)	5.20 (1.10)	3.83 (1.40)
(i) Upgrading of product quality	3.31 (1.37)	3.65 (1.46)	3.67 (1.53)	3.20 (1.16)	2.36 (1.04)	4.29 (0.95)	3.14 (1.17)
(ii) Making for more product innovation	3.44 (1.45)	3.58 (1.48)	3.90 (1.30)	3.33 (1.37)	2.76 (1.48)	4.14 (1.07)	3.43 (1.50)
Group 3 **Inter-firm competition/rivalry**	4.60 (1.67)	5.43 (1.34)	5.38 (1.72)	4.23 (1.23)	3.23 (1.45)	4.29 (0.49)	4.36 (2.17)
Group 4 **Linkages with foreign or domestic firms and institutions**	4.10 (1.37)	4.41 (1.39)	4.95 (1.36)	3.84 (1.07)	3.65 (1.29)	3.86 (1.46)	3.43 (1.40)
(i) Related firms (agglomerative economies)							
(ii) Universities and other research institutions	3.29 (1.30)	3.40 (1.14)	3.52 (1.29)	3.32 (1.40)	3.17 (1.46)	2.83 (1.47)	3.00 (1.30)

Source: Author's estimates, based on the survey of sample firms.

resources (in the survey, these firms were all based in resource-rich countries).[22] Relative to Japanese firms, too, European and United States firms appeared to value the access to foreign-based created assets, e.g.

---

[22] Viz. Australia, Canada and New Zealand.

organizational capacity, management, expertise and relational skills, more highly than firms in other developed countries, although, in general, the competitive advantages of Japanese firms were perceived to be more home-specific than those of their major competitors. (This also reflects their lower degree of transnationality (see sect. 3*c*).)

In respect of each of the other three components of Porter's diamond, European-owned firms appeared to rely more on foreign (than on domestic), sources of competitiveness compared with their United States and Japanese counterparts. This was especially the case with respect to inter-firm competition and, for firms from smaller European countries, with respect to linkages with related firms.

Notwithstanding the below-average significance attached by the sample firms to the acquisition of, or access to, foreign-created assets, the rankings—even for TNCs from large and advanced economies, such as Germany and the United States, which record a comparative patenting advantage in high-technology products (Cantwell and Hodson, 1991)—suggest that large TNCs in the high and medium-technology industries are increasingly seeking a technology-related presence[23] in each of the main Triad countries or regions (e.g. United States firms in the European Union (EU) and Japan; Japanese firms in EU and the United States, and EU firms in Japan and the United States).

(c) *Degree of transnationality*

Perhaps the most popular hypothesis about the propensity of TNCs to derive competitive advantages from their foreign operations is that the competitive advantages will be positively related with the extent and depth of a firm's foreign operations, relative to domestic operations. In this section, two exercises are attempted. The first exercise is to group the data presented earlier on the basis of the four bands of transnationality set out in the introduction. The second exercise is to relate the extent of the two measures of transnationality identified earlier—viz. the average of the percentage of global assets and employment outside the home country and the average

---

[23] By this is meant either the pursuance of R & D-related activities by the foreign affiliates and/or strategic alliances with foreign firms, or the monitoring of the innovatory activities of foreign firms.

Table 6. The sourcing of competitive advantages, classified by extent of transnationality[a] of the sample firms, 1993

(Mean value)

	All	Low	Medium Low	Medium High	High
Group 1 **Access to resources and assets**	3.54 (1.39)	2.33 (1.28)	2.83 (1.13)	3.76 (1.00)	4.50 (1.31)
(i) Natural resources	4.24 (1.70)	3.82 (2.01)	3.90 (1.84)	4.21 (1.34)	4.62 (1.84)
(ii) Unskilled labour	3.98 (1.70)	2.26 (1.58)	3.57 (1.80)	4.38 (1.44)	4.52 (1.60)
(iii) Skilled and professional labour	2.98 (1.27)	2.06 (1.18)	2.48 (0.90)	3.33 (1.19)	3.48 (1.20)
(iv) Innovatory capacity	2.88 (1.40)	2.82 (1.81)	2.67 (1.20)	2.80 (1.32)	3.43 (1.22)
(v) Organizational capacity	3.12 (1.32)	1.94 (1.25)	2.88 (1.26)	3.37 (1.09)	3.57 (1.19)
(vi) Managerial expertise	3.24 (1.42)	2.35 (1.27)	2.71 (1.30)	3.55 (1.29)	3.73 (1.28)
(vii) Relational skills	3.75 (1.62)	2.35 (1.84)	3.21 (1.38)	4.13 (1.29)	4.43 (1.48)
Group 2 **Consumer demand**	3.94 (1.59)	3.27 (1.98)	3.32 (1.39)	3.98 (1.43)	4.96 (1.49)
(i) Upgrading of product quality	3.31 (1.37)	3.00 (1.80)	2.79 (1.28)	3.34 (1.11)	4.18 (1.33)
(ii) Making for more product innovation	3.44 (1.45)	2.94 (1.48)	3.17 (1.46)	3.49 (1.36)	4.04 (1.32)
Group 3 **Inter-firm competition/rivalry**	4.60 (1.67)	4.29 (1.79)	4.08 (1.67)	4.68 (1.56)	5.37 (1.43)
Group 4 **Linkages with foreign or domestic firms and institutions**	4.10 (1.37)	3.06 (1.48)	3.50 (1.25)	4.27 (1.25)	4.87 (1.14)
(i) Related firms (agglomerative economies)					
(ii) Universities and other research institutions	3.29 (1.30)	2.12 (0.93)	2.87 (1.55)	3.54 (1.15)	3.90 (1.11)

Source: Author's estimates, based on the survey of sample firms.

[a] For a more precise definition of the extent of transnationality, see the introduction.

percentage of global R & D expenditures undertaken outside the home country—to the values placed on Porter's four facets of competitive advantages (groups 1 to 4), in a series of bi-variate relationships.

Table 7. Bi-variate correlation coefficients between perceived competitive advantages derived from foreign activities and the extent of transnationality of the sample firms, 1993

(Correlation coefficients)

	A	B
Natural assets (group 1 [i-ii])	0.31982	0.18998
Created assets (technological) (group 1 [iii-v])	0.36478	0.37895
Created assets (managerial) (group 1 [vi-viii])	0.43228	0.39619
Consumer demand (group 2)	0.36339	0.35007
Inter-firm rivalry (group 3)	0.27146	0.28016
Linkages with related firms (group 4 [i])	0.46452	0.39511

Source: Author's estimates, based on the survey of sample firms.

NOTE: A = Extent of transnationality = average foreign employment and assets.
B = Extent of transnationality = average foreign R & D.

Table 6 shows quite clearly that a greater degree of transnationality of firms is likely to be associated with the perception that an increasing proportion of global competitive advantages is derived from foreign sources. This is the most striking finding presented here. It would also appear that, for created assets (group 1 [iii-vii]), the biggest rise in the significance of foreign operations for global competitiveness occurs when the average degree of transnationality (in terms of assets and employment) is 30 per cent or above. However, for innovatory capacity, it is only when the degree of transnationality exceeds 60 per cent that the (average) ranking exceeds 3.00; while, as far as the influence of consumer demand and inter-firm rivalry is concerned, it exceeds 5.00. The distinction between the sources of competitive advantages perceived to be derived from the foreign activities of firms, or from cross-border strategic alliances is particularly noticeable in the case of firms with low and high degrees of transnationality.

Table 7 shows some bi-variate correlation coefficients between the two measures of transnationality (mentioned above) and six indices of competitive advantage. It suggests that, for the main measure of transnationality (A in table 7), the coefficients are positive and significant at the 99 per cent level or above, and that the most likely benefits of increased transnationality are likely to arise from access to foreign organizational capacity and managerial expertise, or linkages forged with foreign firms.

The results for the second measure of transnationality, the foreign R & D ratio (B in table 7), are virtually identical to the results obtained from the first measure. Except for the correlation with natural assets (group 1 [i-ii]), all of the coefficients are positive and significant at the 99 per cent level or above. Not surprisingly, the two measures of transnationality are also significantly positively correlated with each other (0.69, $p<0.0001$), confirming that they essentially measure aspects of the same phenomenon.[24]

(d) *Size of firm*

The final contextual variable which, to some extent, is correlated with the degree of transnationality,[25] is the size of the firm. While it is not obvious why size itself should be related to the geographical sourcing of competitiveness-enhancing assets, the literature (Dunning, 1993, chap. 6) suggests that large firms are more likely to engage in FDI than small firms, and that transnationality itself may help a firm to preserve (or increase) its share of global markets.

At the same time, it might be hypothesized that medium-size firms are likely to be more specialized in their portfolio of global assets and, hence, more reliant on foreign sources to enhance or complement that portfolio. However, in recent years, the evidence suggests that very large firms are just as likely to engage in merger-and-acquisition (M & As) activities as are their smaller competitors. Finally, it might be predicted that the portfolio of foreign competitiveness-enhancing assets sought by medium-size firms,[26] relative to larger firms, may have more to do with gaining access to specialized resources, capabilities and markets, and/or establishing linkages with local foreign firms. (Medium-size firms are typically expected to engage in less competition with foreign-owned firms than their larger counterparts.)

The results of the table 8 survey show that the size of firms is only of marginal importance in affecting the sourcing of most categories of competitive advantage.

---

[24] Since the group 1-4 variables are not continuous, the significance levels reported here should be taken as indicators of a possible relationship (to be investigated further in a future article), rather than a form of final analysis.

[25] Although the Pearson correlation coefficient between size and transnationality is not significantly different from zero.

[26] It should be remembered that all the firms in the sample are large, relative to the universe of firms, but that some are much larger than others.

Table 8. The sourcing of competitive advantages, classified by the size of the sample firms, 1993

(Mean value)

	All	Medium (<5 billion dollars)	Large (5-24.9 billion dollars)	Very Large (25-49.9 billion dollars)	Mega (>50 billion dollars)
Group 1 **Access to resources and assets**	3.54 (1.39)	3.58 (1.54)	3.53 (1.44)	3.57 (1.09)	3.43 (0.94)
(i) Natural resources	4.24 (1.70)	3.96 (1.79)	4.41 (1.76)	4.20 (1.61)	4.69 (1.03)
(ii) Unskilled labour	3.98 (1.70)	4.03 (1.68)	3.89 (1.92)	4.25 (1.39)	3.86 (1.17)
(iii) Skilled and professional labour	2.98 (1.27)	2.96 (1.28)	3.06 (1.41)	2.88 (1.12)	2.86 (0.77)
(iv) Innovatory capacity	2.88 (1.40)	2.71 (1.41)	2.98 (1.49)	3.25 (1.13)	2.71 (1.20)
(v) Organizational capacity	3.12 (1.32)	3.13 (1.39)	3.22 (1.39)	3.00 (1.15)	2.79 (0.89)
(vi) Managerial expertise	3.24 (1.42)	3.45 (1.56)	3.09 (1.38)	3.19 (1.42)	3.14 (0.95)
(vii) Relational skills	3.75 (1.62)	3.82 (1.66)	3.78 (1.72)	3.50 (1.59)	3.64 (1.08)
Group 2 **Consumer demand**	3.94 (1.59)	4.07 (1.68)	3.90 (1.63)	3.93 (1.58)	3.67 (1.15)
(i) Upgrading of product quality	3.31 (1.37)	3.30 (1.38)	3.40 (1.46)	3.13 (1.25)	3.14 (1.17)
(ii) Making for more product innovation	3.44 (1.45)	3.57 (1.46)	3.33 (1.48)	3.53 (1.46)	3.29 (1.33)
Group 3 **Inter-firm competition/rivalry**	4.60 (1.67)	4.58 (1.85)	4.72 (1.60)	4.50 (1.26)	4.29 (1.68)
Group 4 **Linkages with foreign or domestic firms and institutions**	4.10 (1.37)	4.24 (1.36)	4.07 (1.45)	4.25 (1.29)	3.57 (1.16)
(i) Related firms (agglomerative economies)					
(ii) Universities and other research institutions	3.29 (1.30)	3.28 (1.29)	3.30 (1.41)	3.20 (1.08)	3.43 (1.16)

Source: Author's estimates, based on the survey of sample firms.

4. Which foreign countries provide the most valuable assets to upgrading competitiveness?

The next section gives some details of the foreign countries most frequently identified by the sample firms as sources of competitiveness-enhancing assets. Here, it is not possible to evaluate the relative importance of one source country *vis-à-vis* another country; rather, the kinds of competitive advantages that are likely to be most closely associated with each country are identified.

What do the received theories of trade and investment predict in this respect? The answer is that much will depend on the type of FDI undertaken (i.e. market seeking, resource seeking, efficiency seeking, or strategic-asset seeking) and on the static and dynamic competitive advantages of the home and host countries. Also, like trade, FDI and strategic alliances might be either between or within industries. In the case of the former, one might expect the FDI recipient countries to provide assets in which the home country is deficient. In the case of the latter, FDI is likely to be more of the efficiency-enhancing and strategic-asset seeking variety, in which case the foreign country will tend to provide assets similar to those in which the investing firm (or country) has a competitive edge.

In carrying out the survey, each respondent was asked to name up to three foreign countries which were hosts to the firm's foreign affiliates and which, in their perception, have had the most positive impact on their own competitiveness. Table 9 presents some evidence derived by relating each of the competitive advantages identified to the number of times a particular host country was mentioned.

The data in table 9 confirm most of the predictions made here. Thus, developing countries clearly have a competitive advantage in affording access to natural resources, while Germany and Japan provided more than the average share of technological assets, and the United States more than the average share of managerial assets and consumer-demand conditions favouring competitiveness (e.g. large markets). United States firms were also perceived to offer a more competitive stimulus than other firms. Relative to other countries, Germany, Japan and the United States were perceived as good locations for establishing linkages with other firms.

Table 9. The sourcing of competitive advantages, classified by region or country, 1993

(Number of times countries are mentioned among three most important sources of advantages)

	Europe				North America		Rest of the World			
	France	Germany	United Kingdom	Other	United States	Canada	Japan	Developed countries	Developing countries	Sum
Group 1 **Access to ressources** Natural assets	0.50	2.00	5.00	4.00	6.00	2.00	0.50	2.50	30.00	52.50
Created assets (technological)	1.33	12.33	6.67	9.00	18.00	1.33	7.33	1.67	7.67	65.33
Created assets (managerial)	1.00	4.33	6.00	3.67	10.00	2.67	3.33	1.33	6.33	38.66
Group 2 **Consumer demand**	0.00	4.00	4.00	0.00	8.00	0.00	0.00	0.00	3.00	19.00
Group 3 **Inter-firm rivalry**	2.00	4.00	4.00	5.00	16.00	1.00	2.00	1.00	4.00	39.00
Group 4 **Linkages with related firms**	6.00	8.00	7.00	4.00	17.00	1.00	9.00	0.00	4.00	56.00
Totals	10.83	34.66	32.67	25.67	75.00	8.00	22.16	6.50	55.00	270.4
Number of firms	(139)	(131)	(125)	(116)	(113)	(138)	(118)	(142)	(130)	9

Source: Author's estimate, based on the survey of sample firms.
NOTE: The number of firms reflects the sample firms excluding firms from the host country.

One surprising finding of the survey was the frequency with which developing countries were identified as providing access to created assets and to competitive advantages (groups 3 and 4). Although mostly this result is the outcome of pooling a number of countries at different stages of development into the same group, China and Brazil were both mentioned a number of times as providing a locale for more efficient production via economies of scale.[27] As might be expected, the host countries named varied on the basis of both the industry of the investing TNCs and their home countries. For example, firms in HT sectors accounted for 45-57 per cent of all mentions in the high-technology countries, i.e. the United States, Germany, Japan and the United Kingdom. Taking European, United States and Japanese firms as a group, 25 per cent of the countries mentioned by the respondents as sources of competitive advantages were located elsewhere in the Triad (or, alternatively, nearly three quarters of their advantages were derived from within the Triad).

5. Modes of foreign entry most likely to advance competitive advantages

The respondents of the sample firms were asked to rank (on a scale of 1 to 7) the importance of each of three modes of acquiring and/or tapping into the resources and capabilities of foreign countries, i.e. the "foreign" diamonds of competitive advantage. The modes are FDI, non-equity cooperative agreements (e.g. strategic alliances, management contracts, licensing and franchising agreements etc.) and arm's length transactions (in both intermediate and final goods and services).

The hypothesis here is that deeper forms of cross-border structural integration, such as FDI and non-equity cooperative agreements, are more likely to result in an addition of competitive advantages to the home company (and country) than shallower forms of transactions, such as arm's length trade. It might also be predicted that firms are more likely to internalize their assets in industries that are technologically intensive than in those that are not. A related hypothesis is that non-equity cooperative ventures are likely to be ranked higher as a modality for competitive advantages in less

---

[27] The figures in table 9 somewhat understate the importance of scale economies, as the figure for group 2 indicates the overall importance of consumer demand conditions and not scale economies specifically.

Table 10. Perceptions of the importance of the mode of foreign involvement of the sample firms, by technological intensity, extent of transnationality and size, 1993

(Mean value)

	Foreign direct investment	Non-equity arrangements	Arm's length transactions
All firms	5.22 (1.63)	4.66 (1.46)	3.96 (1.45)
High technology	5.23 (1.63)	5.31 (1.08)	4.27 (1.59)
Medium technology	5.53 (1.42)	4.62 (1.41)	3.55 (1.28)
Low technology	5.02 (1.74)	4.07 (1.57)	3.92 (1.36)
Low transnationality	4.06 (2.19)	4.71 (1.69)	3.59 (1.62)
Medium-low transnationality	5.46 (1.67)	5.25 (1.07)	4.25 (1.26)
Medium-high transnationality	5.44 (1.38)	4.83 (1.28)	4.12 (1.53)
High transnationality	5.28 (1.67)	3.93 (1.65)	3.55 (1.21)
Medium size	4.83 (1.84)	4.48 (1.50)	4.00 (1.41)
Large	5.19 (1.61)	4.69 (1.31)	3.95 (1.46)
Very Large	6.07 (0.80)	5.07 (1.79)	4.13 (1.46)
Mega	5.93 (0.92)	4.79 (1.58)	3.71 (1.64)

Source: Author's estimates, based on the survey of sample firms.

technological-intensive industries than other modes of foreign entry. Because of the need to integrate their global operations, it might be expected that these firms that are the most transnationalized may be expected to rank FDI higher than firms whose foreign operations are less significant to their overall prosperity. Finally, it might be supposed that firms from countries whose domestic institutions are organized along hierarchical lines (e.g. the United States) would be more likely to rank FDI as a foreign asset-acquiring modality than firms from countries whose institutions practice more cooperative modes of governance (e.g. Japan).

Tables 10 and 11 offer evidence to assess these hypotheses. Overall, it is clearly true that deeper forms of integration were perceived to offer the most benefits as regards the acquisition of competitive advantages abroad. In table 10, the average rankings of each of the three modalities were classified according to the technological intensity of the firm. While the LT and MT firms behaved as expected, in other words, FDI was perceived to offer the most advantages, and arm's length transactions were perceived to offer the least, HT firms claimed that they derived more advantages from cooperative alliances than from FDI.

Table 11. Perceptions of the importance of the mode of foreign involvement of the sample firms, by region or country of origin, 1993

(Mean value)

	Foreign direct investment	Non-equity arrangements	Arm's length transactions
All firms	5.22 (1.63)	4.66 (1.46)	3.96 (1.45)
"Large" European countries	5.47 (1.42)	4.60 (1.45)	3.81 (1.24)
"Small" European countries	4.70 (1.95)	4.81 (1.63)	4.32 (1.34)
United States	4.97 (1.80)	4.58 (1.26)	3.29 (1.64)
Japan	5.96 (1.17)	5.36 (1.41)	4.88 (1.26)
Developing countries	4.79 (1.37)	4.29 (1.14)	4.21 (1.25)
Other countries	4.57 (2.15)	3.14 (1.57)	3.29 (1.38)

Source: Author's estimates, based on the survey of sample firms.

Also, somewhat contrary to what was expected, the medium-low and medium-high transnationalized firms actually derived more advantages from all three modalities of foreign involvement than did firms in the high-transnationalization group. Whether this is another manifestation of the law of diminishing returns, or of the differences between the first mover versus late mover experiences, remains to be seen. More in line with what was expected, the connection between internationalization and size of firm was confirmed. The figures in table 10 appear to indicate that the gains from FDI accrue predominantly to the largest firms, whereas the benefits from cooperative alliances and trade do not exhibit any obvious pattern.

Finally, table 11 sheds some light on the advantages derived from the different modalities classified by the nationality of the firm. The results here contradict the hypothesis that firms from countries with typically hierarchical organizational structures would prefer FDI over other modes of foreign entry. Japanese firms stand out not only in terms of their perceived advantages being derived from FDI, but also in terms of their perceived advantages being derived from cooperative ventures and arm's length trade. In summary, although FDI is most likely to be the preferred route by which the domestic and foreign diamonds of competitive advantages are linked, the relative significance of this route is the greatest in the case of (large) firms that have medium-to-high technology and are moderately-to-highly transnational.

Table 12. Perceptions of firms as to whether foreign sourcing of competitive advantages has become relatively more important in recent years, 1993

(Mean value)

All	1.50 (1.36)	All	1.50 (1.36)
High technology	1.69 (1.35)	Medium size	1.47 (1.32)
Medium technology	1.79 (0.95)	Large	1.31 (1.55)
Low technology	1.16 (1.51)	Very Large	1.87 (0.83)
"Large" European countries	1.64 (1.27)	Mega	2.00 (0.96)
"Small" European countries	1.90 (1.04)	Low transnationality	0.94 (1.55)
United States	1.26 (1.53)	Medium-low transnationality	1.29 (1.37)
Japan	1.62 (1.39)	Medium-high transnationality	1.62 (1.44)
Developing countries	1.43 (0.85)	High transnationality	1.96 (0.92)
Other countries	0.38 (2.00)		

Source: Author's estimates, based on the survey of sample firms.

6. The dynamics of the geographical sourcing of competitive advantage

To what extent do firms perceive that their access to foreign diamonds of competitive advantage is becoming a more important contribution to their overall competitive advantages? It might be hypothesized that, as firms increase their degrees of transnationality, they will augment increasingly—rather than exploit—their competitive advantages from their foreign operations.[28] In the course of the survey, firms were asked to rank, on a scale of −3 to +3, whether they perceived that their competitive advantages were becoming less or more dependent on their foreign operations (indicated by negative or positive signs).

The results presented in table 12 essentially confirm the impressions already gained from tables 10 and 11 concerning the benefits derived from internationalization. Although all the sample firms, overall, perceived that the foreign sourcing of competitive advantages has become more important over time, medium- and high-technology firms had a more positive view of that importance, as did the very large and mega firms. When classified by their regions of origin, firms from small European countries were the most positive in their responses, a finding that may not only reflect their size, but also the aftermath of the wave of investment created by the perceived threat of a "fortress" Europe by 1992 for the countries outside the European Union. Not surprisingly, the firms that were the most transnationalized also perceived the influence of foreign sourcing of competitive advantages to be very important.

---

[28] For an interesting article that supports this contention in relation to the R & D activity of TNCs in the electronics and pharmaceuticals industries, see Kuemmerle (1996).

7. Home-country governments and competitiveness

The role of governments in affecting the competitiveness of firms is a subject which has long been of fascination to economists and other scholars (Dunning, 1997, forthcoming). In the present context, we are interested in the perceived influences of home governments on the ability and willingness of the sample firms to be competitive globally. In the questionnaire, eighteen possible ways in which governments might exert such an influence were identified. For each of these ways, firms were asked to assign a figure from a range of –3 to +3, according to whether or not, in their opinion, home governments had influenced their *global* competitiveness negatively or positively between 1988 and 1993 (0 would indicate no influence at all). Table 13 presents the findings classified by the three groups of industries used here. Table 14 does the same for firms classified on the basis of nationality and table 15 does the same for firms classified on the basis of the degree of transnationality.

Table 13 shows that, on average, the sample firms thought that actions by their home governments either had a marginally beneficial effect on their global competitiveness (e.g. the provision of infrastructure, trade, or industrial and technology policies), or a marginally adverse effect (e.g. in terms of market facilitating, social and environmental policies). There were few consistent differences between industry groupings, except that LT firms thought more favourably of most government actions (notably, trade, industrial, technology, monetary and exchange rate policies) than either HT or MT firms.

More interesting differences are revealed in table 14. While United States and European firms perceived that their home governments pursued

Table 13. Perceived influence of home governments on competitive advantages of the sample firms, by technological intensity, 1993

(*Mean value*)

		All	High technology	Medium technology	Low technology
1.	Provision and upgrading of infrastructure	0.52 (1.00)	0.48 (1.08)	0.60 (1.00)	0.51 (0.93)
2.	Social policies	-0.13 (1.44)	-0.10 (1.47)	-0.31 (1.42)	-0.05 (1.43)
3.	Monetary and exchange-rate policies	-0.11 (2.16)	-0.04 (1.94)	-0.73 (2.00)	0.20 (2.40)
4.	Trade policies	0.35 (1.45)	0.26 (1.40)	-0.12 (1.52)	0.70 (1.39)
5.	Industrial and technology policies	0.36 (2.08)	0.35 (2.11)	0.06 (2.46)	0.54 (1.80)
6.	Education and training policies	0.09 (1.25)	0.15 (1.43)	0.03 (1.29)	0.07 (1.04)
7.	Environmental policies	-0.06 (1.24)	-0.34 (1.22)	0.09 (1.33)	0.13 (1.16)
8.	Market-facilitating policies	-0.37 (1.99)	-0.38 (2.05)	-0.55 (2.15)	-0.25 (1.85)
9.	Promoting an ethos of competitiveness	0.28 (2.91)	0.53 (3.57)	0.13 (2.72)	0.13 (2.30)
10.	Promoting a culture of investment and saving	0.18 (1.91)	0.51 (2.12)	-0.09 (2.07)	0.04 (1.56)

Source: Author's estimates, based on the survey of sample firms.

macroeconomic, i.e. monetary and exchange rate and trade policies marginally favourable to their competitiveness, Japanese firms considered the same policies of their home government to work to their disadvantage. By contrast, Japanese firms believed that most macro-organizational policies of their Government (e.g. environmental, industrial and technology policies), together with the promotion of a culture of investment and saving, aided their global competitiveness, while United States respondents generally thought their Government worked to the detriment of their competitiveness in terms of these policies. Respondents of firms from countries with small and/or relatively unsophisticated domestic markets were generally more favourably disposed to the actions of their home governments than those of

Table 14. Perceived influence of home governments on competitive advantages of the sample firms, by region or country of origin, 1993

(*Mean value*)

		All countries	"Large" European countries	"Small" European countries	United States	Japan	Other developed countries	Developing countries
1.	Provision and upgrading of infrastructure	0.52	0.68	0.81	-0.08	0.44	0.67	1.00
		(1.00)	(1.03)	(0.96)	(0.85)	(0.96)	(0.67)	(0.99)
2.	Social policies	-0.13	-0.45	0.38	-0.74	0.20	-0.13	0.92
		(1.44)	(1.50)	(1.50)	(1.37)	(1.00)	(1.55)	(1.12)
3.	Monetary and exchange-rate policies	-0.11	-0.21	0.52	0.29	-1.62	-0.13	-0.08
		(2.16)	(2.33)	(2.46)	(1.44)	(1.77)	(1.81)	(2.43)
4.	Trade policies	0.35	0.30	0.29	0.19	-0.19	1.13	1.54
		(1.45)	(1.15)	(1.95)	(1.49)	(1.47)	(0.64)	(0.88)
5.	Industrial and technology policies	0.36	-0.12	1.19	-0.53	0.72	0.13	2.08
		(2.08)	(1.87)	(2.40)	(1.96)	(2.03)	(1.46)	(1.38)
6.	Education and training policies	0.09	0.07	0.81	-0.58	0.12	0.00	0.62
		(1.25)	(1.32)	(1.47)	(0.81)	(0.95)	(0.93)	(1.50)
7.	Environmental policies	-0.06	-0.14	0.00	-0.55	0.50	-0.25	0.31
		(1.24)	(1.32)	(1.00)	(1.36)	(0.86)	(1.39)	(1.18)
8.	Market-facilitating policies	-0.37	-0.37	-0.14	-1.13	-0.54	-0.25	1.31
		(1.99)	(2.02)	(2.31)	(1.87)	(1.79)	(1.39)	(1.38)
9.	Promoting an ethos of competitiveness	0.28	0.21	0.57	0.10	0.19	-0.50	1.17
		(2.91)	(3.00)	(3.65)	(3.28)	(1.30)	(2.20)	(3.41)
10.	Promoting a culture of investment and saving	0.18	0.29	0.38	-0.77	0.84	-0.50	0.85
		(1.91)	(2.00)	(1.72)	(1.81)	(1.70)	(1.07)	(2.19)

Source: Author's estimates, based on the survey of sample firms.

firms from other countries. Rather surprisingly, while firms from all the major investing countries believed that their governments did little to facilitate the efficient workings of markets, firms from developing countries were noticeably more appreciative in this respect.

Table 15 suggests that the degree of transnationality of firms seems to be unrelated to their perception of the competitiveness-enhancing actions by their home-country governments. Thus, while the least transnationalized firms regarded industrial and technology policies and the promotion of an investment and savings culture more favourably than did the most transna-

Table 15. Perceived influence of home governments on competitive advantages of the sample firms, by extent of transnationality, 1993

(Mean value)

		All firms	Low firms	Medium-low firms	Medium-high firms	High firms
1.	Provision and upgrading of infrastructure	0.52 (1.00)	0.67 (1.12)	0.67 (0.87)	0.37 (1.03)	0.61 (0.87)
2.	Social policies	-0.13 (1.44)	-0.39 (1.20)	0.10 (1.64)	-0.27 (1.39)	0.03 (1.59)
3.	Monetary and exchange-rate policies	-0.11 (2.16)	-0.39 (2.35)	-0.68 (1.89)	-0.03 (2.09)	0.83 (2.19)
4.	Trade policies	0.35 (1.45)	0.50 (1.62)	0.32 (1.59)	0.27 (1.41)	0.60 (1.38)
5.	Industrial and technology policies	0.36 (2.08)	1.33 (1.97)	0.59 (2.17)	0.17 (1.96)	-0.13 (2.13)
6.	Education and training policies	0.09 (1.25)	0.28 (1.49)	-0.09 (1.06)	0.17 (1.15)	-0.10 (1.45)
7.	Environmental policies	-0.06 (1.24)	-0.06 (1.43)	0.00 (0.98)	-0.08 (1.39)	-0.07 (1.08)
8.	Market-facilitating policies	-0.37 (1.99)	-0.06 (1.66)	-0.41 (1.44)	-0.53 (2.18)	-0.40 (2.03)
9.	Promoting an ethos of competitiveness	0.28 (2.91)	0.24 (2.25)	0.32 (2.93)	0.33 (2.79)	0.13 (3.73)
10.	Promoting a culture of investment and saving	0.18 (1.91)	0.72 (1.84)	0.23 (2.43)	-0.02 (1.59)	0.03 (2.18)

Source: Author's estimates, based on the survey of sample firms.

tionalized firms, the reverse appeared to be the case for macroeconomic and social policies. Here, however, bi-variant comparisons tend to break down. It so happens that Japanese firms are the least transnationalized, and it could be (and most likely is) that the country of the respondent firms is a more important determinant of the role of government than the degree of their transnationality. It is precisely for this and related reasons why a multi-variant approach is necessary to complement the descriptive interpretation of the data set out here.

8. Summary and some policy conclusions

The findings of the survey show that a not insignificant part of the competitive advantages of some of the world's leading industrial TNCs are derived from their foreign-based activities. It has also been revealed here that a firm's competitive advantages are dependent on the types of FDI and/or cross-border alliances in which the firm engages; the technology intensity of the industries in which the firm is involved; the country from which the firm emanates; and, particularly, the degree of the firm's transnationality. The findings of the survey also reveal that an overwhelming majority of firms from all industries and countries believed that the importance of the foreign sourcing of their competitive advantages had increased in recent years, and that FDI (followed by inter-firm cooperative agreements) is the favoured modality for acquiring these advantages.

The role of home-country governments in influencing the global competitive advantages of the sample firms is also discussed briefly. In general, that role was not perceived to be of critical significance. However, there were noticeable differences observed among firms of different nationalities, and between large developed countries (particularly the United States and Asian developing countries, especially Japan, and smaller European countries, such as Sweden) that *inter alia* reflect some very distinctive country-specific philosophies of the role of governments, hierarchies and markets.

The policy implications of the findings are straightforward. Governments need to recognize that firms engage in foreign activities both to exploit their existing competitive advantages and to protect, or augment, these advantages. There is also a strong suggestion that, as the foreign operations of firms become a more significant component of the world economy, technology, organizational skills and other tangible assets are likely to be transferred across national boundaries, not only from the investing firms to their affiliates abroad, but also from the foreign affiliates back to the investing firms. This is particularly likely to be the case for TNCs (e.g. Asea Brown Boveri, Unilever, IBM) that operate from multiple home bases. This being so, it behoves governments to pursue as liberal policies towards both inward and outward FDI as their macro-organizational strategies would allow; at the same time, governments can pursue policies that help create the kinds of domestic economic environments that enable indigenous firms to become strong contestants in the global marketplace, and the country to attract high value-added FDI into its domestic arena. ■